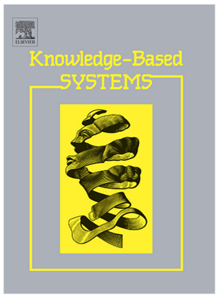
Accepted Manuscript



Group recommender systems: A multi-agent solution

Christian Villavicencio, Silvia Schiaffino, J. Andres Diaz-Pace, Ariel Monteserin

PII: S0950-7051(18)30557-4
DOI: https://doi.org/10.1016/j.knosys.2018.11.013
Reference: KNOSYS 4576

To appear in: *Knowledge-Based Systems*

Received date : 2 May 2018
Revised date : 9 November 2018
Accepted date : 12 November 2018

Please cite this article as: C. Villavicencio, S. Schiaffino, J. Andres Diaz-Pace et al., Group recommender systems: A multi-agent solution, *Knowledge-Based Systems* (2018), https://doi.org/10.1016/j.knosys.2018.11.013

Research highlights

- A group recommender approach based on multi-agent systems is proposed.
- The approach replaces the traditional aggregation techniques with negotiation.
- The group members are satisfied in a more even way than with traditional approaches.

Group Recommender Systems: A multi-agent Solution

Christian Villavicencio[a], Silvia Schiaffino[a], J. Andres Diaz-Pace[a], Ariel Monteserin[a]

[a]*ISISTAN (CONICET - UNCPBA), Campus Universitario, Tandil, 7000, Buenos Aires, Argentina*

Abstract

Providing recommendations to groups of users has become a promising research area, since many items tend to be consumed by groups of people. Various techniques have been developed aiming at making recommendations to a group as a whole. Most works use aggregation techniques to combine preferences, recommendations or profiles. However, satisfying all group members in an even way still remains as a challenge. To deal with this problem, we propose an extension of a multi-agent approach based on negotiation techniques for group recommendation. In the approach, we use the multilateral Monotonic Concession Protocol (MCP) to combine individual recommendations into a group recommendation. In this work, we extend the MCP protocol to allow users to personalize the behavior of the agents. This extension was evaluated in two different domains (movies and points of interest) with satisfactory results. We compared our approach against different baselines, namely: a preference aggregation algorithm, a recommendation aggregation algorithm, and a simple one-step negotiation. The results show evidence that, when using our negotiation approach, users in the groups are more uniformly satisfied than with traditional aggregation approaches.

Keywords: recommender systems, group recommendations, multi-agent systems, negotiation

1. Introduction

Nowadays, when a user wants to purchase a product, contract a service or do some activity (e.g., watching a movie), she often faces the problem of information overload [1, 2]. This is because users must deal with a variety of potentially interesting items in the target domain. In this context, a Recommender System (RS) allows users to identify those items that match their needs, preferences,

Email addresses: christian.villavicencio@isistan.unicen.edu.ar (Christian Villavicencio), silvia.schiaffino@isistan.unicen.edu.ar (Silvia Schiaffino), andres.diazpace@isistan.unicen.edu.ar (J. Andres Diaz-Pace), ariel.monteserin@isistan.unicen.edu.ar (Ariel Monteserin)

tastes, and goals. To carry out this task, several recommendation techniques have been proposed in the literature [3, 4].

Most RS techniques have been developed to assist individual users. However, in some domains such as movies, music or tourism, the recommendations can serve groups of people as well as individuals. A Group Recommender System (GRS) looks for item recommendations that are good for a group of users as a whole. That is, the recommended items should satisfy, as much as possible, the individual preferences of all group members [5]. Generating recommendations that satisfy a group of users with possible competing interests is not straightforward. *Traditional approaches* make use of aggregation techniques in order to produce group recommendations [6, 7, 8].

Although these traditional approaches are used in several domains, they still present some limitations. First, aggregation techniques can produce values that might not represent correctly the data being aggregated, especially when the data to be aggregated is small and has a high variance. For example, let us assume that 3 users rate a movie with scores 1, 1 and 5, respectively; then the (aggregated) group rating can be: 2.33 (if the *average* aggregation technique is used), 1 (if *misery minimization* or *plurality voting* is used), and 5 (if *most pleasure* is used). From this simple example, we can see that neither of those aggregated ratings truly represent the ratings given by the users. Because of this problem, recommendations generated with aggregation techniques seldom satisfy the group members in a uniform way. Second, the decision-making process of the group and the group dynamics [9, 10] are not reflected by the aggregation technique [6, 9, 8]. By dynamics, we mean any kind of group behavior during the decision making process that can influence the decision result, such as: users' concession profiles, users' tolerance or influence, among others. As a result, traditional approaches can produce a recommendation that does not match the group's interests or neglects one of the group members to satisfy others, which often leads to the whole group rejecting the recommendation.

When a human group has to choose an item, its members generally discuss and analyze their options for achieving a consensus on the item. Different methods can be used to achieve consensus, such as: voting, auction-biding or negotiation [11]. Particularly, the negotiation method is expected to generate recommendations that satisfy the group members more uniformly than traditional approaches. Along this line, we proposed a multi-agent approach for group recommendation called PUMAS-GR [12], which was later re-named to MAGReS (*Multi-Agent Group Recommender System*). In MAGReS, a personal agent represents each user of a group. This personal agent knows the user's preferences and acts on her behalf when making item proposals to other agents and looking for agreements. These agreements are achieved by the agents through a cooperative negotiation process. Particularly, we chose a multilateral negotiation method known as Monotonic Concession Protocol (MCP) [13], since it closely mirrors the way in which human negotiation seems to work [14]. We argue that this approach helps to increase the quality of the group recommendations by increasing the satisfaction of the group as a whole.

The general idea of the approach was initially presented in [12], including preliminary evaluation results. However, the agents were all equipped with one single acceptance criterion for item proposals, which did not allow agent personalization by the users. By personalization, we mean that a user can change the behavior of her representative agent to better capture her beliefs and how the agent should act (accordingly) during a group negotiation. Thus, we argue that different acceptance criteria need to be supported, so that each agent can decide whether to accept or reject a given item proposal. Furthermore, in some domains it is important to consider recommendations that contain items already rated by group members (in the past). These criteria can help to produce recommendations being closer to what users expect, thus increasing the probability of those recommendations being accepted by the group. In this context, the main contribution of this article is the development of two new decisions strategies for the agents in MAGReS, which are also empirically assessed. We additionally analyze the amount of information that a given user needs to reveal (to other agents/users) when making group recommendations.

To demonstrate the feasibility of MAGReS with the two proposed strategies, we have evaluated MAGReS in the movies domain (MovieLens) and in the points of interest (POI) recommendation domain (Yelp). We performed different experiments, and compared our proposal against traditional approaches for recommendation aggregation, preference aggregation, and a simple one-step negotiation. The comparisons were carried out using both item-based and user-based recommender systems. The effectiveness of the group recommendations was assessed in terms of the average satisfaction of a group and how uniformly the group members were satisfied. The results showed that the average group satisfaction reported was high for the recommendations produced by MAGReS, and also the group members were satisfied in a more uniform way than with traditional appraoches.

The rest of the article is organized into 4 sections as follows. In Section 2 we discuss related works. In Section 3 we present the details of the MAGReS approach. Then, in Section 4 we report on the experiments carried out to evaluate our approach. Finally, in Section 5 we give the conclusions and outline future lines of work.

2. Related Work

The generation of group recommendations began to be investigated in the RS field during the last decade [7, 15]. Traditional GRS can be classified into three main categories according to the strategy they follow to generate the recommendations. These categories include: (i) those that perform "recommendation aggregation" [6], by producing individual recommendations for every member of the group and then merging those recommendations; (ii) those that perform "preference aggregation" [16, 17], by merging the individuals' preferences/ratings in order to obtain a group evaluation for each candidate item; and (iii) those that perform "model/profile aggregation" [18], by merging individuals' models into a single group model first and then generating suggestions based on that model. There are various techniques for aggregating data, and their suitability depends on both the data being aggregated (it is not possible, for example, to aggregate non-numeric data with an average) and the goal being pursued. When it comes to merging individual recommendations, Masthoff [10] analyzed different techniques such as: average, average without misery, and least misery, among others. In [6], the authors analyzed the effectiveness of ranked list recommendations tailored to a group of users using different methods such as: Spearman footrule, Borda count, average and least misery [7, 10].

Multi-agent systems (MAS) have been applied in several areas and domains (see Chapter 10 of [14]). Regarding RS, some approaches have proposed MAS to generate recommendations both for individuals and groups. The way in which RS and MAS are combined depends on the main goal of the combination. In some cases, they are combined as a way to improve the quality of the recommendations [21, 22, 23]. In other cases, a MAS is used for recommendations because its architecture allows developers to model the problem more adequately and prescribes a clear assignment of responsibilities to the different system modules [24, 25, 26, 27, 28, 29, 30]. Both ideas have been used together in [31, 32, 33]. There are many references in the literature about MAS being applied to RS for individual users in domains such as: adaptive customization of websites [28], tourism [24, 26], games on mobile phones [34], TV shows [31], training courses [27], and e-commerce [32, 23], among others.

However, when it comes to MAS being applied to GRS, only a few works have been reported, mostly in the tourism domain. In [21], the authors present a GRS in the tourism domain that relies on the application of cooperative agent-based negotiation. The agents act on behalf of group members and participate in a direct (alternating offers) or mediated (merging rankings) negotiation, which ultimately produces group recommendations based on individual recommendations and user preference models. One of the limitations of this approach is that it has only been tested with simulations involving two agents (therefore, it is a group of two people). We argue that this kind of approaches should be assessed in bigger groups in order to be practical for RS.

In [25] an agent-based negotiation schema that uses alternating offers is developed, in which the agents (each one representing a group member) negotiate the preferences of the whole group. This approach is said to be suitable for every domain, provided that the domain can be represented using ontologies. Like MAGReS, this approach does not use aggregation techniques, but a negotiation process in order to compute the group preferences. In [35] a refined version of the approach is presented, in which new agent types are introduced for filtering the list of items according to the group preferences (resulting from the negotiation) and for mediating the negotiations among user agents. Our work differs from [25, 35] in that they negotiate user preferences while MAGReS negotiates recommendations. A similar idea is used by [36], with agents representing group members and a mediator to coordinate the agents' work. An interesting aspect of [36] is that a user agent not only tries to protect the interests of its represented user, but also models her behavior with respect to conflicting situations. In this way, the attitude and behavior of the agent changes (and adapts to the situation) when looking for agreements during the negotiation. However, there are two differences between MAGReS and [36], namely: (*i*) a protocol based on a *Merging Ranks* technique (instead of MCP), and (*ii*) a preference aggregation techniques used by the mediator agent (when computing the group rating for each item).

In [33] a MAS-based system called e-Tourism is presented, which is able to produce both individual and group recommendations. Nonetheless, this system differs in that the agents are used for representing users but also for modeling components of the RS. This approach differs from ours in that it relies on aggregation techniques to generate the group recommendations (specifically, profile aggregation techniques are used in order to aggregate the group members' profiles into a single group profile).

In [37] the authors present a review of the state of the art in RS using MAS from a game-theory point of view, with the objective of introducing examples of how social-choice mechanisms can be extended using social information extracted from the analysis of interactions within a social network.

In [10] a GRS for points of interest (POI) that uses dynamic elicitation of user preferences is proposed. The approach allows users to iteratively express and revise their preferences during the decision making process through a chat-based app . Although the main idea of this approach is similar to our approach, , the app requires the users to discuss about the options and vote them, while in MAGReS the discussion and negotiation are carried out by the agents.

3. Proposed Multi-agent Approach

In order to address the problems discussed in Section 1, we initially proposed a multi-agent negotiation approach called PUMAS-GR (later renamed to MAGReS) as an alternative to aggregation techniques. In this article, we extend MAGReS and develop two new strategies for the approach (see Section 3.4 and Section 3.5), so as to enrich the modeling of the users' criteria and improve both the quality and the "acceptability" of the recommendations. Additionally, in Section 3.6 we discuss how the proposed approach affects the information privacy of the group members, in terms of how much information from each user needs to be revealed to the rest of the group members (during the negotiation).

3.1. MAGReS

MAGReS (*Multi-Agent Group Recommender System*) is based on a MAS in which each agent acts on behalf of a group member. Each agent maintains a profile that contains the user's preferences and it is capable of (i) predicting the rating the user would assign to an item not yet rated, and (ii) generating a ranking of "interesting items" for the user (items the user would like). Initially, the user's preferences are the ratings assigned by the user to the items she rated in the past. MAGReS differs from traditional approaches (see Figure 1) that normally work with a central entity, which employs some aggregation technique (either for user profiles, user preferences or single-user recommendations) in tandem with a single-user recommender system. The novelty of MAGReS is the

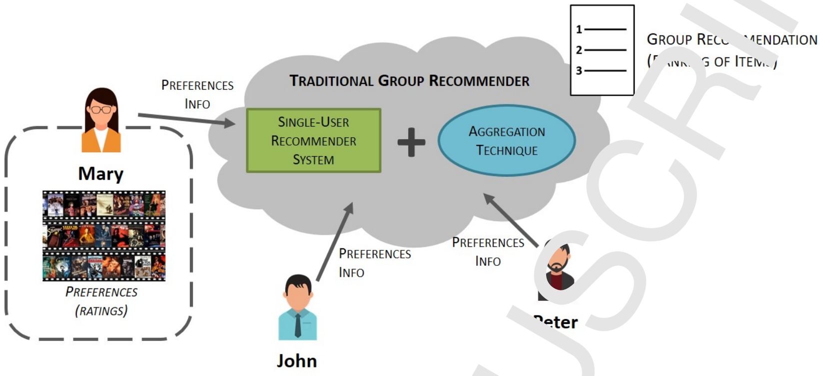


Figure 1: Traditional approach to group recommendation

replacement of aggregation techniques by a negotiation process in which a group of *User Agents* (i.e., agents that represent the users) try to reach a consensus on the most satisfying items for the group. Although various negotiation protocols are possible, only a few of them have addressed two important properties for us, namely: (i) mimic the negotiation process followed by humans, and (ii) be suitable for multi-lateral negotiation. Based on these considerations, we chose the protocol known as MCP (*Monotonic Concession Protocol*) [13] to guide the negotiation. A general overview of our approach is shown in Figure 2. More

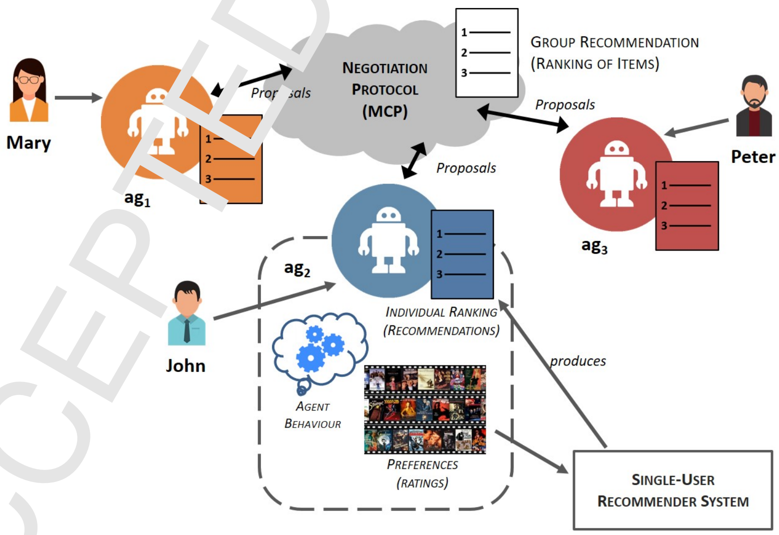


Figure 2: Proposed approach

Formally, let $A = \{ag_1, ag_2, ..., ag_n\}$ be a finite set of N cooperative agents, and let $X = \{x_1, x_2, ...x_m\}$ be a finite set of potential agreements or proposals, each one of them containing an item that can be recommended to one of the agents.

Each agent $ag_i \in A$ has a *utility function* $U_i : X \rightarrow [0,1]$ that maps proposals to its satisfaction value. In our approach, each agent internally relies on a Single-User RS (SUR) to generate a ranking containing the items (candidate proposals) proposed by the agent. The ranking is sorted in descendant order according to the utility value of the item. This way, the set X can be seen as the union of the rankings produced for all the agents, plus an special agreement called *conflict deal*, which yields utility 0 for all the agents and will be chosen as the worst possible outcome (no agreement is possible).

For example, let us assume we want to generate recommendations of movies to (groups of) users. Along this line, let us have a group of three friends who want to watch a movie together, and a set of M possible movies to be chosen. According to MAGReS, each user is equipped with her own personal agent that is able to access her user profile. For simplicity, a profile includes only ratings over (a subset of) the possible movies. A user rating $rt_i(item)$ is a value (in the range $[0,1]$ where 0 means dissatisfaction and 1 means high satisfaction) assigned by the user i to the given item (i.e., a movie). Additionally, the utility function of each agent $ag_i \in A$ is defined as follows:

$$U_i(x_j) = \begin{cases} rt_i(x_j) & if\, x_j \in R_i \\ SUR_i(x_j) & if\, x_j \notin R_i \end{cases} \tag{1}$$

Where R_i is the list of items rated by user i (represented by ag_i) and $SUR_i(x_j)$ is the rating predicted by the SUR_i, which is the single-user recommender system[1] used by ag_i for generating its list of candidate proposals.

In this context, let us consider the following (initial) situation:

- ag_1 handles ratings $< rt_1(M5) = 0.75, rt_1(M3) = 0.56) >$ on behalf of user #1,
- ag_2 handles $< rt_2(M10) = 0.82, rt_2(M52) = 0.65 >$ for user #2, and
- ag_3 handles $< rt_3(M32) = 0.88, rt_3(M46) = 0.8 >$ for user #3.

Then we have $A = \{ag_1, ag_2, ag_3\}$ as the MAS in which the negotiation for the "best" movie (the one that will satisfy all the three agents) takes place.

3.2. Monotonic Concession Protocol (MCP)

The agents engage in rounds of negotiation, each one making proposals (of items) that need to be assessed by the other agents, until an agreement is reached or the negotiation finishes with a *conflict*. Our agents abide by a set of predefined rules that specify the range of "legal" moves available at each agent at any stage of the negotiation process. In MCP, these rules have to do with:

[1] Most single-user recommender systems provide a way to predict the rating of a user for a given item. For example, the Mahout framework provides the *estimatePreference* function that accepts the IDs of both a user and an item as parameters and returns the predicted rating that particular user would assign to the item.

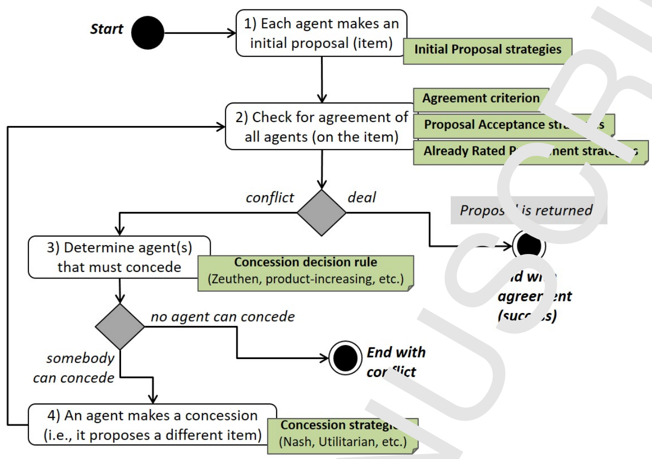


Figure 3: Steps of the MCP for MAGReS (adapted from [13])

(i) the agreement criterion, (ii) which agent makes the next concession (after a round with no agreement), and (iii) how much an agent should concede. The protocol presents some limitations, namely: (i) an agent cannot influence the negotiation position of other agents (e.g., by exchanging a justification), and (ii) an agent needs to assign quantitative utilities to proposals (often, via a utility function). Additionally, in contrast to the benefits provided by the multi-agent system, there are some issues that might affect the performance of the recommender system. For example, negotiation involves reasoning (i.e. computation cost) and communication (i.e. communication overhead) [38]. However, we focus on two properties of the selected negotiation protocol to diminish the effects of these issues: termination and deadlock-freedom. The MCP protocol guarantees (a) that any negotiation process following the protocol will eventually terminate (termination), and (b) that at least one agent can concede satisfying the concession criterion at any negotiation stage, until an agreement has been reached. Taking into account both properties, as the overall number of agents is finite, there can only be a finite number of concessions. Thus, the negotiation is bound to terminate reaching an agreement or finishing in conflict. The steps of the MCP protocol are summarized in Figure 3. At the beginning (step 1), each agent makes an initial proposal according to its own *Initial Proposal* strategy (for example, the agent proposes its "favorite" or top-ranked item if its *Initial Proposal* strategy is the one called *Egocentric*, i.e. it always selects the proposal with the highest utility value for the agent). Then, the initial proposals of all the agents are exchanged in order to determine if an agreement over one of the proposals can be reached. The notion of *agreement* (also called deal) is defined in terms of the utility of a given proposal for the agents. Thus, there is an agreement if one agent makes a proposal that is at least as good (regarding utility) for any other agent as their own current proposals. Formally, we rely on the following criterion:

> *Multilateral Agreement Criterion*[2]*:* An agreement is reached if and only if there is an agent $ag_i \in A$ such that its proposal x_i is accepted by every other agent $ag_k \in A$. Whether a proposal is accepted by each agent $ag_k \in A$ is determined by the *Proposal Acceptance* (PrA) strategy used by that agent (see Section 3.5).

If an agreement is reached, the proposal that satisfies all the agents is chosen (if several proposals meet this criterion, we simply pick one of them randomly). On the contrary, if no agreement can be reached, one (or more) of the agents must concede. A concession means that an agent seeks an inferior proposal (for example, in terms of its own utility), with the hope of reaching an agreement. If none of the agents can concede, the process finishes with no-agreement (the *conflict deal* is returned). Several concession strategies are possible (see Section 3.3). Also, note that even though an agent can exclude from its proposals those items that her user has already rated, certain items proposed by the agent might have been rated by some other user within the group. Along this line, users' opinions with regard to recommendations of items they have already rated vary from one user to another. Given that agents should take this situation into account when assessing proposals sent by other agents, we propose the *Already-Rated Punishment* (ARP) strategies for MCP (see Section 3.4).

3.3. Concession strategies

As it was explained before, when no agreement is reached during a round of negotiation, at least one of the agents must concede. At this point, there are two problems that need to be solved, namely: (i) determining which agent(s) has to make the concession and (ii) selecting the next item to propose (i.e. the new proposal).

3.3.1. Concession Decision Rule

From all the agents that can make an effective concession (i.e. those having alternative proposals in their pool of proposals), it is necessary to find the one(s) that must concede in the next round of negotiation. One way for selecting the conceding agent(s) is to apply the Zeuthen strategy [39] around the concept of *willingness to risk conflict* (WRC). This strategy was initially designed to be used in bilateral negotiations and later extended to work in multi-lateral ones, as explained in [12][3]. The WRC for ag_i (WRC_i) is then given by Equation 2:

$$WRC_i = \begin{cases} 1 & if\, U_i(x_i) = 0 \\ \frac{U_i(x_i) - min\{U_i(x_k) | k \in A\}}{U_i(x_i)} & otherwise \end{cases} \tag{2}$$

[2] If the *Strict PrA* strategy is used this criterion works as the one proposed in [13]

[3] The "product-increasing" and "sum of products" strategies proposed by Zeuthen were also generalized (from bilateral to multilateral negotiations) in [13] and are valid within our framework, but we did not use them in the experiments reported in this article.

Where $U_i(x_i)$ is the utility value for ag_i over the item it proposed (x_i is the most recent proposal made by ag_i), and $U_i(x_k)$ is the utility value for ag_i over the item the agent ag_k proposed (x_k). Once the WRC_i of every agent is computed, the agent(s) with the lowest WRC_i value must make a concession. In case two or more agents hold the same WRC value, different strategies can be followed. Without losing generality, we make all those agents concede in our implementation framework.

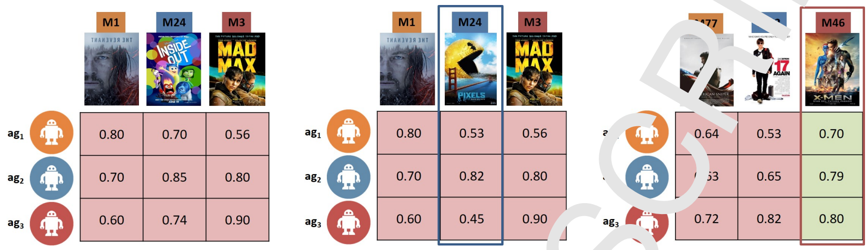
3.3.2. Concession strategy

Various strategies are discussed in the literature for deciding on the item the conceding agent(s) should propose in the next round [13]. For our work, we initially selected the so-called *Nash concession* strategy, which states that an agent makes a proposal such that the product of utilities of the other agents increases (*Nash product*). When assessing the behavior of the Nash strategy in practical cases, we observed that sometimes it leads to "early conflicts" during the negotiation. An early conflict is a situation in which all agents quickly exhaust their potential proposals and the MCP ends with no deal. This is because once the *Nash product* is high enough, the amount of candidate proposals reduces drastically, and so does the number of agents that can make concessions. This behavior therefore drives the negotiation to a premature end, and also makes the agents discard potential agreements just because their Nash product was not high enough. The same behavior was observed when using the *Utilitarian concession* strategy (it uses addition instead of product of utilities). To mitigate this problem, we defined a variation of the Utilitarian concession strategy called *Desires Distance.*

Desires Distance (DD). DD attempts to measure how far a candidate proposal is with respect to the desires of the other agents. Along this line, an agent makes a proposal that is "closer to the other agents' desires" (we denote the desires distance as dd_{value}) but also has a utility value lower or equal than the agent current proposal. DD guarantees termination and deadlock-freedom, the demonstration follows from that for Utilitarian concession [13].

The DD strategy works as follows. Initially, we consider that agent ag_i must make a concession and therefore find a new item to propose (in the next round of negotiation). To do this, we create a list with all the "eligible" candidate proposals (i.e., proposals with a lower utility value than the current proposal), and then we select the first candidate from the list whose dd_{value} is lower or equal to the one of the current proposal. The dd_{value} is computed as explained in Equation 3. This strategy requires $U_k(x_i) - U_k(x_k) < 0$, because otherwise agent ag_k is already satisfied and therefore it should not be considered in the distance computation.

$$dd_{value}(x_i) = \sum_{k=0, k\neq i}^{N} \mid U_k(x_i) - U_k(x_k) \mid$$
$$where\; U_k(x_i) - U_k(x_k) < 0 \; and \; x_k \; is \; the \; current \; proposal \; of \; ag_k \quad (3)$$



(a) Round 1 (b) Round 2 (c) Round k

Figure 4: MCP negotiation example over movies

Figure 4 shows an example of MCP when having a group of three users, each one represented by one of the three agents (ag_1, ag_2 and ag_3). In the first round (Figure 4a) each agent proposes a movie: ag_1 proposes *M1*, ag_2 proposes *M24* and ag_3 proposes *M3*. According to the previously defined *agreement criterion* (Figure 3, step 2) there is no agreement, as none of the proposals satisfies all the agents and, therefore, one (or more) of the agents has to make a concession. The *concession decision rule* determines that ag_2 must concede. For this example, we used *WRC* (see Equation 4). At the beginning of *Round 2* (Figure 4b), ag_2 uses DD to select her next proposal and as a result she proposes *M10* (as $dd_{value}(M10) = 0.26$, $dd_{value}(M24) = 0.23$ and $dd_{value}(M10) \leqslant dd_{value}(M24)$). All the proposals are assessed again and the *agreement criterion* determines that ag_1 and ag_3 will reject *M10*. The process repeats until reaching *Round k*, in which ag_3 proposes the movie *M46* that satisfies all the agents and therefore the negotiation ends successfully. Movie *M46* will be then recommended to the group of users. There are two observations that can be inferred from the example. The first one is that the negotiation is not guaranteed to terminate successfully (with an agreement) among all the agents. The second one is that more than one agreement are possible, meaning that in round k an agreement could have existed over more than one proposal. In such a case, one of the agreements should be selected and several selection strategies can be followed (e.g. random selection, choose the one that maximizes the satisfaction of the agents, among other criteria).

3.4. Already-Rated Punishment strategy

The recommended items will determine whether the recommendation is (partially or fully) accepted by the group. Depending on the approach for generating the recommendations, items that were rated by some of the group members in the past might be recommended (again). For example, if a traditional preference-aggregation approach is used, during the preference aggregation stage, the RS needs to determine which items will be part of the group preferences profile (i.e. those items whose ratings are computed using the aggregation technique). The approach for selecting items might vary from one RS to another. If all the items rated by at least one of the group members are used during the aggregation process, then there is no way an item being already rated by a group member will be included in the recommendation. However, if only a subset of the items rated by at least one of the group members is considered, those items that were not included in the group preference profile have a chance of being recommended, especially if they received a high rating by the users.

During the field test of our approach [40], we observed that while some users would give a low feedback to recommendations containing items they had already rated, others gave a high feedback to items they had liked in the past, and others simply stated that they would take the item again if their friends did not take it. This means that not every user might reject recommendations containing items already rated or always accept them. Therefore, we believe that agents should take this aspect into account during the negotiation.

The *Already-Rated Punishment* (*ARP*) strategy is proposed as a way to model the user behavior explained above. *ARP* works as follows: when an agent ag_i is asked by another agent ag_k about its utility regarding a proposal x_k (made by ag_k), ag_i first uses its utility function to assess x_k, and then applies the *ARP* strategy to compute a penalty to be applied before the utility is reported. This way, *ARP* can indirectly influence ag_k decision with respect to its next proposal (assuming that ag_k concession strategy is not egocentric, i.e. it cares about other agents' opinions). The penalty is also taken into account when the agent decides whether to accept a proposal, and thus proposals that received a higher penalty will be less likely to be accepted. Depending on the *ARP* strategy variant used by ag_i (which is determined by its associated user), the utility reported to ag_k will vary. At the moment MAGReS supports five variants for the *ARP* strategy, but other variants can be included in the future. These five variants are the following:

- Easy-Going: No penalty is applied (the penalty is zero), the agent does not care about receiving proposals with an item already rated by its user.
- Flexible: The penalty is computed as $penalty = 1 - flexibilityLevel$, where *flexibilityLevel* (f) is a value between 0 and 1 and it models how flexible the user is with regard to receiving recommendations of items she has already rated. The higher the flexibility is, the lower the penalty should be. For example: if $flexibilityLevel = 0.75$ and the rating given by the user[4] is 0.8, the penalty is computed as $penalty = 1 - 0.75 = 0.25$ therefore the reported utility is $0.8 - 0.25 = 0.55$.
- Min-Satisfaction: The penalty is 0 (zero) if the rating given by the user exceeds the *minSatisfaction* (ms) value set[5], otherwise the penalty is the whole utility value which would result on the agent reporting an utility of 0 for the proposal (meaning that she would not accept a recommendation

[4] According to Equation 2 the utility is the same as the rating given by the user.

[5] Both the *flexibilityLevel* and the *minSatisfaction* of the variants *Flexible, Min-Satisfaction* and *Flexible Plus* can be either set explicitly by the user or learnt by using Machine Learning techniques.

with that item). For example, let us suppose *minSatisfaction* = 0.6, then if the rating given by the user is 0.5 the reported utility will be 0, but if the rating given is 0.65 then the agent will report 0.65 as the utility value of the proposal.

- Flexible Plus: It is a combination of the *Flexible* and *Min-Satisfaction* strategies. Firstly, it computes the penalty in the same way the *Flexible* strategy does. Then, if the rating given by the user surpasses the *minSatisfaction* value set, the penalty is divided by two, otherwise the penalty is not modified.
- Taboo: As the users prefer not to receive recommendations with items already rated, the penalty for proposals containing those items will be equal to the utility value defined by the agent's utility function (according to Equation 2, it corresponds to the rating given by the user to the item), and therefore, the utility reported by the agent for those proposals will be 0.

Depending on the domain, some variants could produce better results than others. For example, the *Easy-Going* variant could be useful in the music and restaurant recommendation domains, as users do not tend to reject recommendations containing items they have already consumed. The *Min-Satisfaction* variant could be useful in other domains, like books, food, movies and music, where users would not mind to consume again an item they liked. Finally, the *Taboo* variant can be useful in mostly all domains, but especially in those where the activity of consuming an item (e.g., watching a movie, or traveling to a certain location) requires a significant amount of effort, time or resources (e.g., money).

3.5. Proposals Acceptance Strategy

Not all the users follow the same criteria (nor use the same strategy) when deciding whether to accept a proposal from others, and the same is valid for the agents. Depending on the criteria used by the users (group members), one proposal can be acceptable or not. This aspect was not taken into account in our previous work, where the *Multilateral Agreement Criterion* determined that a proposal is accepted by an agent if it is better than its current one. While for some users a proposal might not be acceptable because it is not strictly better than their current proposals; for others, the proposal might be good enough and therefore acceptable, although not being strictly better. This way, if all the agents were to use the same acceptance strategy, many potential agreements would be discarded. Moreover, even if all the agents were to use the same strategy, a strict strategy might not be always ideal to model their associated users.

Let us come back to the example presented in Figure 4 and suppose that in between rounds 2 (Figure 4b) and k (Figure 4c) there is a round i depicted by Figure 5. In such a case, according to the *Multilateral Agreement Criterion*, no agreement was reached. Interestingly, if we look closely, one of the proposals

(*M71*) could be considered as good enough to have an agreement because it is better than what ag_1 and ag_2 proposed and it is almost as good as the current proposal of ag_3. If all the agents use a strict criterion, *M71* is not considered as an agreement, the negotiation will carry on and will later end with an agreement on *M46* (see Figure 4c), which does not satisfy the users as much as *M71* does. This situation might be different if each agent, and particularly ag_3, had its own acceptance criteria. Not only the negotiation would have finished earlier, but also the quality of the recommendation would have been higher (because the satisfaction of each group member would have increased). Along this line, we propose the *Proposal Acceptance* (*PrA*) strategy as a way to introduce a user decision criterion into the agent model.

The *PrA* strategy currently has three variants, but others can be defined so as to improve the model. These three variants are the following:

- Strict: The standard definition of multilateral agreement is used[6]. An agent ag_i will accept a proposal x_k if it is at least as good (in terms of utility value) as its own proposal (see Equation 4).

$$accept(ag_i, x_k) = true \quad iff \quad U_i(x_k) \geq U_i(x_i) \tag{4}$$

- Relaxed: An agent accepts a proposal x_k if it is as good as its own proposal or very close to what she wants. Equation 5 depicts the formalization of this strategy, where *relaxPercentage* (*rp*) can assume values in the $[0, 0.2]$ range[7]. Each agent might have a different *rp*, which is determined by the represented user, depending on her own criteria[8].

$$accept(ag_i, x_k) = true \quad iff \quad U_i(x_k) \geq U_i(x_i) * (1 - relaxPercentage) \tag{5}$$

- Next: A proposal x_k is accepted if it is better than the agent's next proposal (i.e. the one it would make if it had to concede in the next round). This strategy is similar to the one used by the agents in the single-issue negotiation model proposed in [41], in which an agent accepts an offer if its utility is higher than that of the agent counter-offer. It works under the following assumption: "if I reject proposal x_k, which is not better than my current proposal x_i but better than my next one (x_{i+1}), and as a result the round ends with no agreement, I am forced to make a concession and my next proposal is accepted, then I would have lost some utility because I rejected proposal x_k that was better than my next one" (see Equation 6).

$$accept(ag_i, x_k) = true \quad iff \quad U_i(x_k) \geq U_i(x_{i+1}) \tag{6}$$
$$where\; x_{i+1}\; is\; ag_i\; next\; proposal$$

[6]This is the one used by all agents in [12].
[7]The upper limit of the range was determined empirically.
[8]This value can be set explicitly by the user, or learnt automatically by using Machine Learning techniques over user feedback, among other options.

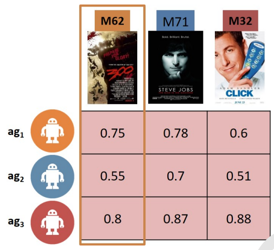


Figure 5: Example of potential agreement, discarded due to a strict criterion (Round i, with 2<i<k).

3.6. Information privacy

Not every user will be comfortable with her private information being shared, made public or exposed in any way [5]. In the recommendation area, the amount and type of information that every user needs to reveal about herself, so as to allow the recommender system to produce a recommendation for her, depends on the approach used to make the recommendations. For example, when using a traditional approach based on preference aggregation, the system will require the user to inform about all her preferences. This is another advantage of using a multi-agent approach like MAGReS: the agents can control the amount of information that needs to be revealed during the recommendation process. For example, the only information about the user preferences being revealed is the one related to: (i) the proposals the agent makes, and (ii) the utility values she reports when asked if she is willing to accept a proposal made by another agent. This way, MAGReS is expected to be more respectful of the information privacy (i.e. by leaking less users' private information) than a traditional approach.

4. Evaluation

In this section we describe the experiments carried out to evaluate our approach for group recommendation in two domains: movies and points of interest (POI). In Section 4.1 we describe the datasets used. In Section 4.2 we establish the algorithms considered as baselines. In Section 4.3 we define the evaluation criteria. In Section 4.4 we describe the objectives of our experiments and explain some aspects of their design, such as the parameters taken into account and their values. At last, in Section 4.5 we discuss the results obtained.

4.1. Datasets

Two datasets were used to evaluate our approach: the first one in the movies domain and the second one in the POI domain. In both cases, we randomly sampled 180 users (without repetition) and created 45 groups of 3, 4 and 5 people (15 groups per size). The two datasets are described below.

- Movies: we used the popular *MovieLens*[9] dataset with ratings of users for different movies to generate user groups of varying sizes. In particular, the experiments were performed with *MovieLensLatestSmall* that contains 100,000 ratings, 700 users and 9,000 movies. We will refer to this dataset as ML_LATEST_SMALL.
- POI: we used a dataset from the Yelp Dataset Challenge[10] that contains check-ins (visits of users to places) from various cities in the world. For the purpose of the experiments we only considered those check-ins belonging to Arizona (US), since the resulting network is dense. Also, we only considered those users having at least 9 check-ins. The resulting dataset contains 19,193 users. Regarding the distribution of check-ins, the dataset contains 497,029 check-ins in 85,901 POIs generated by the selected users. We will refer to this dataset as YELP.

4.2. Description of baselines

We compared the recommendations resulting from traditional group recommendation against those produced by MAGReS. As the baseline for this comparison, we implemented two recommender systems that rely on traditional approaches, namely:

- TRADGRec-PA: a GRS that uses preference aggregation, as proposed in [20, 42].
- TRADGRec-RA: a GRS that relies on the aggregation of recommendations produced for each group member. It is based on the approach proposed in [42](chapter 2, section 2.4), which generates a recommendation containing k items for each group member and then merges all those recommendations into a single recommendation (the group recommendation). For this matter, the group preference (or rating) of each candidate (i.e., an item recommended to a group member) is computed by aggregating the preferences (either existing or predicted ones) of each group member over the candidate, and then the k candidates with the highest group preference values are selected to be part of the group recommendation.

For MAGReS, we also implemented a protocol known as the *One-Step protocol* [43] as an additional baseline. The *One-Step protocol* is a variant of MCP in which all the negotiations happen in one single round. The agents simply interchange their proposals (one proposal each) and seek for an agreement, but there is no concession.

The single-user recommender (SUR) system used by the agents in MAGReS, TRADGRec-PA and TRADGRec-RA, was implemented using the Duine

[9] http://grouplens.org/datasets/movielens/

[10] https://www.yelp.com/dataset_challenge

framework[11] and the Mahout framework[12]. Duine was used in the first implementation stages and helped us to gain confidence that the approach was viable [44, 12]. Nonetheless, Duine had some limitations regarding the loading time of the data models, which made it unsuitable for a responsive recommendation application[13]. Therefore, all the tests in this article were made using the Mahout-based implementation, which uses a CF-based (Collaborative Filtering) approach. In the framework, both item-centered CF (ICF) and user-centered CF (UCF) filtering are supported, and thus we implemented two variants of the SUR, one using ICF and another one using UCF.

4.3. Evaluation criteria

4.3.1. Satisfaction

The satisfaction of the recommendations for the users were measured in terms of several indicators. These indicators can be computed both at the item- (i.e., an item recommended) and recommendation- (i.e., a list of items recommended) levels. These indicators are the following:

- **group satisfaction** (S_g **or GS**): it measures the satisfaction of the group with respect to an item (or a list of items) being recommended. The satisfaction is equivalent to the estimated preference (or rating) of the user/group with respect to the item.
  - item level: the group satisfaction for an item x_j is computed as explained in Equation 9, where n is the number of group members ($\mid g \mid$) in the group (g) and $S_i(x_j)$ is the satisfaction of group member u_i over item x_j. $S_i(x_j)$ is computed as the rating for the pair $< u_i, x_j >$ predicted by the SUR by using the rating prediction CF formula[14] and the similarity metric chosen for each experiment.

$$GS(x_j) = S_g(x_j) = \frac{\sum_{i=0}^{n} S_i(x_j)}{n} \tag{9}$$

  - recommendation level: the GS of a recommendation r consisting of k items ($r =< x, x_2, ..., x_k >$) is computed as the average of the GS

[11] http://duineframework.org/

[12] http://mahout.apache.org/

[13] For example, Duine needed around 30 seconds to load the data models when using the MovieLens100K dataset, while Mahout only needed 2 seconds to perform the same task.

[14] The way a rating r_{u_i,x_j} is predicted depends on whether the SUR uses ICF (see Equation 7) or UCF (see Equation 8).

$$r_{u_i,x_j} = \bar{r}_{x_j} + \frac{\sum (r_{u_i,x_k} - \bar{r}_{x_k}) \times Similarity(x_j, x_k)}{\sum \mid Similarity(x_j, x_k) \mid} \tag{7}$$

$$r_{u_i,x_j} = \bar{r}_{u_i} + \frac{\sum (r_{u_k,x_j} - \bar{r}_{u_k}) \times Similarity(u_i, u_k)}{\sum \mid Similarity(u_i, u_k) \mid} \tag{8}$$

of every item in r (see Equation 10).

$$GS(r) = S_g(r) = \frac{\sum_{j=0}^{k} GS(x_j)}{k} \tag{10}$$

- **members satisfaction dispersion (MSD)**: it assesses how uniformly the group members are satisfied by either a single item x_j or a recommendation r. The lower the MSD is the more uniformly satisfied the group members will be.
  - item level: as it can be seen in Equation 11, the MSD for an item x_j is computed as the standard deviation of the group members satisfaction.

$$MSD(x_j) = \sqrt{\frac{\sum_{i=0}^{n}(S_i(x_j) - S_g(x_j))^2}{n}} \tag{11}$$

  - recommendation level: the MSD for a recommendation r that consists of k items ($r =< x_1, x_2, ..., x_k >$) is computed as the average of the MSD for every item in r (see Equation 12).

$$MSD(r) = \frac{\sum_{j=0}^{k} MSD(x_j)}{k} \tag{12}$$

- **fairness**: it is a metric proposed in [45] for evaluating a recommendation of an item (x_j) to a group. It is defined as the percentage of group members satisfied by the recommendation (see Equation 13). To determine which users are satisfied, the authors set a threshold th to 3.5 stars (out of 5 stars, the equivalent to 0.7 out of 1) and any group member with a satisfaction value above the threshold is considered satisfied. We kept $th = 0.7$ and extended this metric for applying it to a recommendation r of k items. As it can be seen in Equation 14, the fairness of a recommendation r of k items is computed as the average of the fairness of the items.

$$fairness(g, x_j) = \frac{\mid \bigcup_{u_i \in g} : S_i(x_j) > th \mid}{n} \tag{13}$$

$$fairness(g, r) = \frac{\sum_{j=1}^{k}(fairness(g, x_j))}{k} \tag{14}$$

4.3.2. Information Privacy

Another factor that needs to be taken into consideration is the amount of user information that gets leaked during the recommendation process. This information is mainly related to: (i) the utility function held by the user (i.e. the way she computes the ratings for the items) and (ii) the items she can propose during a negotiation (i.e. her candidate proposals). This way, given two recommendation approaches R_1 and R_2, we consider that R_1 produces better recommendations (than R_2) in terms of information privacy if it leaks less information while producing the recommendations.

To measure the amount of information revealed by *UserAgents* of MAGReS we computed two , namely: *UFIL* and *PIL*.

- **Utility Function Information Leak (*UFIL*):** it measures (in a [0; 1] range) the amount of information revealed with regard to the user utility function of user u_i, either when recommending a single item x_j or a list of items $r =< x_1, x_2, ..., x_k >$.

  (i) MAGReS: in this case *UFIL* measures the amount of information related to the utility function that is revealed by agent ag_i (which represents user u_i). At the item-level, we use Equation 15 for agent a_i recommending item x_j. For a list of items (or recommendation), we use Equation 16.

  $$UFIL(u_i, x_j) = UFIL(ag_i, x_j) = \frac{\mid itemsWithUtilityRevealed(ag_i, x_j) \mid}{\mid itemsTotal(ag_i) \mid} \tag{15}$$

  where $\mid itemsWithUtilityRevealed(ag_i, x_j) \mid$ is the amount of items for whom ag_i has revealed its utility (or satisfaction) value when item x_j was recommended; and $\mid itemsTotal(ag_i) \mid$ is the total amount of items over which ag_i can reveal some utility-related information .

  $$UFIL(u_i, r) = \frac{\sum_{j=0}^{k} UFIL(u_i, x_j)}{k} \tag{16}$$

  (ii) TRADGRec-PA: given that TRADGRec-PA uses a traditional approach based on preference aggregation, it requires the system to know everything related to the user preferences. Thus, $UFIL(u_i, x_j)$ is always set to 1.

  (iii) TRADGRec-RA: it uses a traditional approach based on recommendation aggregation where the candidate items are selected by using preference aggregation (the k candidates with the highest group preference value are selected). Thus, it requires the system to know everything related to the user preferences. Again, $UFIL(u_i, x_j)$ is always set to 1.

- **Proposals Information Leak (PIL, MAGReS only):** it represents the proportion (measured in a [0; 1] range) of candidate proposals that the agent (ag_i) revealed during the negotiation process with regard to all the proposals it could have made. We use Equation 17 for a single item x_j (i.e., item level), and Equation 18 for a list of items (or recommendation) $r =< x_1, x_2, ..., x_k >$ (i.e., recommendation level).

$$PIL(u_i, x_j) = PIL(ag_i, x_j) = \frac{\mid candidateProposalsRevealed(ag_i, x_j) \mid}{\mid candidateProposalsTotal(ag_i) \mid} \tag{17}$$

$$PIL(u_i, r) = \frac{\sum_{j=0}^{k} PIL(u_i, x_j)}{k} \tag{18}$$

It is important to notice that the PIL indicator cannot be computed for TRADGRec-PA and TRADGRec-RA, because the concept of proposal does not exist in those systems[15].

4.4. Experimental setting

The number of recommendations for a ranking was set to $k = 3$, 5 and 10 items, since they are common amounts of recommendations (top-three, top-five and top-ten). In the case of MCP (either one-step or multi-step variants), we run the protocol k times, in order to produce the k recommendations. For a given run, we removed from the negotiation space those items that were agreed by all the agents in the previous run.

For all the approaches we conducted experiments with several configurations using both user and item-based recommendation techniques (and different similarity metrics for each one). Mahout allows to use user-based (UB) and item-based (IB) SURs. As it can be seen in Appendix A.1, depending on the type of SUR, different similarity functions are available. In summary, we exercised the three approaches below:

- TRADGRec-PA: it has a single parameter, which is the preference aggregation strategy used for computing the group preferences (aggregated preferences) when building the group preference profile. We used five aggregation strategies: average (AVG), least misery (LM), most pleasure (MP), approval voting (AV) and upward leveling (UL) [7, 20, 42, 46]. The parameters (if any) of each aggregation strategy are discussed in Appendix A.3.1.

- TRADGRec-RA: it has a single parameter that can be configured, which is the aggregation strategy used for computing the group preferences (aggregated preferences) during the process of selecting the candidate recommended items . We used the same four aggregation strategies as in TRADGRec-PA.

- MAGReS: we tested with both the *One-Step* and *MCP* protocols. For each of them, we tested three *Concession* strategies (*Desires Distance*, *Nash* and *Utilitarian*), the five *ARP strategies* (see Section 3.4) and all the *Proposal Acceptance* strategies (see Section 3.5). Despite the variety of parameters that can be tuned, based on the tests we performed, none of them produced a significant impact on the resulting recommendation nor on its quality.

It is important to note that for all the tests of Sections 4.5.1, 4.5.2, 4.5.3, and 4.5.4 we set the amount of items to be recommended (k) up to 10 (i.e., $k = 10$). Once the results of the tests showed the effectiveness of our approach when

[15] For the sake of comparisons, we could assume $PIL = 1$ for the same reasons we set $UFIL$ to 1, but we decided not to compare MAGReS against TRADGRec-PA and TRADGRec-RA with respect to the PIL.

making recommendations of 10 items, we conducted tests with a lower amount of items. A summary of the results for the experiments with $k = 3$ and $k = 5$ can be found in Section 4.5.5.

4.5. Experiments

Given the variety of parameters that can be tweaked within the approach, we decided to report the analysis over those that would have a high impact on both the recommendation process and the recommendations. These parameters are the following:

1. The type of SUR, and especially the type of similarity metric.
   We tested both types of SUR, the user-based and the item-based ones. For the user-based SUR, we conducted experiments with the following similarity metrics: *City Block, Euclidean Distance, Log Likelihood, Pearson Correlation* and *Uncentered Cosine*. For the item-based SUR, we conducted tests with the *City Block, Euclidean Distance, Pearson Correlation* and *Uncentered Cosine*.
2. The *PrA* strategy. In this case we conducted experiments testing all the variants of the *PrA* strategy: *Strict, Relaxed* and *Next*.
3. The *Already Rated Punishment* strategy. In this case we run tests using all the variants of the *ARP* strategy. Particularly, for the *Flexible*, *Min-Satisfaction* and *Flexible Plus* variants, we run tests with different parameterizations.

More details about the parameterization of each approach and strategy can be found in Appendix A.2.2.

4.5.1. Single-User Recommender: User and Item-based

The SUR was one of the factors that affected the most both the recommendation process and the recommendations. During the experimentation, we observed some peculiar results for certain combinations of SUR and similarity metrics:

- When using a user-based (UB) SUR, the *City Block, Euclidean Distance*, and *Uncentered Cosine* metrics caused all the recommendations to be almost the same for both datasets (MovieLens and Yelp) and the 45 groups (per dataset) involved in the test, both for the traditional (TRADGRec-PA and TRADGRec-RA) and MAS-based approaches. According to the results of the tests performed, we believe that the low density of the rating matrices for the datasets and the size of the neighborhoods selected for the experiments might have been the reasons for those results, as they directly affect the effectiveness of most user-based similarity metrics. We analyzed this issue empirically, but a deeper analysis and tests with larger datasets is subject of future work.

- When using a item-based (IB) SUR, the improvement observed for MAGReS in terms of GS, MSD and fairness, was barely noticeable if the *City*

GROUP SATISFACTION (higher is better)	DATASET: ML_LATEST_SMALL							
SIMILARITY METRIC	MAGReS [One-Step; M19]	MAGReS [MCP Strict; M5]	MAGReS [MCP Relaxed; M17]	MAGReS [MCP Next; M11]	TRADGRec-PA [AVG; T1]	TRADGRec-PA [UL; T5]	TRADGRec-RA [AVG; T1]	TRADGRec-RA [UL; T5]
(IB) CITY BLOCK	0.7254 ± 0.0529	0.6297 ± 0.2547	0.7282 ± 0.0465	0.7303 ± 0.0499	0.7263 ± 0.0479	0.7255 ± 0.0476	**[illegible] ± 0.0516**	**0.7395 ± 0.0516**
(IB) EUCLIDEAN WEIGHTED	0.782 ± 0.0726	0.8103 ± 0.182	**0.8472 ± 0.044**	0.8207 ± 0.0689	0.5827 ± 0.1505	0.5718 ± 0.1359	0.7 ± 0.0967	0.7 ± 0.0967
(IB) PEARSON WEIGHTED	0.7752 ± 0.0595	0.9519 ± 0.0387	**0.9545 ± 0.0318**	0.9403 ± 0.0529	0.6069 ± 0.0864	0.6135 ± 0.0938	0.7275 ± 0.0562	0.7275 ± 0.0562
(IB) UNCENTERED COSINE WEIGHTED	0.7782 ± 0.0716	0.8256 ± 0.1336	**0.8464 ± 0.0452**	0.8141 ± 0.0748	0.5667 ± 0.1546	0.5576 ± [illegible]	0.7017 ± 0.0936	0.7017 ± 0.0936
(UB) LOG LIKELYHOOD	0.9309 ± 0.0433	0.9672 ± 0.0113	**0.9681 ± 0.0073**	0.9415 ± 0.0352	0.8655 ± 0.0558	0.8655 ± 0.0558	0.8988 ± 0.0714	0.8988 ± 0.0714
(UB) PEARSON WEIGHTED	0.7336 ± 0.0792	0.9206 ± 0.0378	**0.9371 ± 0.019**	0.8753 ± 0.0383	[illegible] ± 0.0922	0.5898 ± 0.1013	0.6784 ± 0.0755	0.6784 ± 0.0755

GROUP SATISFACTION (higher is better)	DATASET: YELP							
SIMILARITY METRIC	MAGReS [One-Step; M19]	MAGReS [MCP Strict; M5]	MAGReS [MCP Relaxed; M17]	MAGReS [MCP Next; M11]	TRADGRec-PA [AVG; T1]	TRADGRec-PA [UL; T5]	TRADGRec-RA [AVG; T1]	TRADGRec-RA [UL; T5]
(IB) CITY BLOCK	0.7886 ± 0.0724	0.3902 ± 0.4078	0.7885 ± 0.0719	[illegible]	**[illegible]949 ± 0.0662**	0.7908 ± 0.0667	0.7845 ± 0.077	0.7845 ± 0.077
(IB) EUCLIDEAN WEIGHTED	0.671 ± 0.1261	0.9057 ± 0.0488	**0.9116 ± 0.0422**	[illegible] ± 0.0608	0.1854 ± 0.0983	0.1901 ± 0.1182	0.6013 ± 0.0906	0.6013 ± 0.0906
(IB) PEARSON WEIGHTED	0.7208 ± 0.1012	**0.9773 ± 0.0249**	0.9765 ± 0.0219	[illegible]	0.3184 ± 0.1053	0.3277 ± 0.1236	0.6076 ± 0.0777	0.6076 ± 0.0777
(IB) UNCENTERED COSINE WEIGHTED	0.6728 ± 0.1174	0.9064 ± 0.0475	**[illegible]**	[illegible] ± 0.0602	0.1902 ± 0.108	0.1915 ± 0.1137	0.601 ± 0.0905	0.601 ± 0.0905
(UB) LOG LIKELYHOOD	0.7241 ± 0.1094	0.9989 ± 0.0064	**0.9943 ± [illegible]0075**	0.9896 ± 0.0293	0.3881 ± 0.0992	0.3881 ± 0.0992	0.6562 ± 0.1013	0.6677 ± 0.082
(UB) PEARSON WEIGHTED	0.8192 ± 0.0913	0.9926 ± 0.0497	**[illegible]**	0.9757 ± 0.069	0.7225 ± 0.0935	0.7255 ± 0.1129	0.7747 ± 0.0784	0.7747 ± 0.0784

Figure 6: Average group satisfaction (GS) per similarity metric and approach

Block similarity metric was used. The opposite effect was observed when another similarity metric (like *Uncentered Cosine* or *Pearson* Correlation) was used.

The parameterization used for this experiment is specified in Appendix A.4.1. As it can be seen in Figures 6, 7 and 8, the recommendations may vary depending on the similarity metric and type of SUR being used. In general we can see that MAGReS running with the MCP protocol outperformed both the traditional approaches (TRADGRec-PA and TRADGRec-RA) and MAGReS running the One-Step protocol. One exception to the rule was the similarity metric *(IB) City Block*, in which the combination of the *Taboo* variant of the *ARP* strategy and the *Strict* variant of the *PrA* strategy was the reason behind many recommendations being empty. This situation generated a group satisfaction of zero for all the groups affected (6 in the ML_LATEST_SMALL dataset and [illegible]3 in the YELP dataset) and had a direct impact on the values reported in Figures 6 and 7.

The results were validated through a statistical analysis. For each dataset and for each similarity metric, we first performed a Shappiro-Wilk test to determine if the samples were normal. Given that in most cases there was at least one sample that did not follow the normal distribution, we performed a pairwise Wilcoxon Signed Ranks test (which is a non-parametric test) so as to compare each pair of recommendation techniques (for example: *MAGReS [One Step;*

MEMBERS SATISFACTION DISPERSION (lower is better)						DATASET: ML_LATEST_SMALL			
		MAGReS [One-Step; M19]	MAGReS [MCP Strict; M5]	MAGReS [MCP Relaxed; M17]	MAGReS [MCP Next; M11]	TRADGRec-PA [AVG; T1]	TRADGRec-PA [UL; T5]	TRADGRec-RA [AVG; T1]	TRADGRec-RA [UL; T5]
SIMILARITY METRIC	(IB) CITY BLOCK	0.0807 ± 0.0471	**0.069 ± 0.0503**	0.075 ± 0.0406	0.0788 ± 0.0418	0.0754 ± 0.04	[illegible]76 ± [illegible]	0.0801 ± 0.0411	0.0801 ± 0.0411
	(IB) EUCLIDEAN WEIGHTED	0.1121 ± 0.0648	**0.0745 ± 0.0517**	0.0879 ± 0.0498	0.112 ± 0.0831	0.2671 ± 0.1267	[illegible].2496 ± 0.1357	[illegible].1961 ± 0.1045	0.1961 ± 0.1045
	(IB) PEARSON WEIGHTED	0.1053 ± 0.0431	0.0395 ± 0.0306	**0.0357 ± 0.0242**	0.0519 ± 0.0418	0.1502 [illegible] 0.0707	0.1504 ± 0.0683	0.116 ± 0.0493	0.116 ± 0.0493
	(IB) UNCENTERED COSINE WEIGHTED	0.1176 ± 0.0703	**0.0803 ± 0.0487**	0.0847 ± 0.0522	0.1127 ± 0.0935	0.2729 ± [illegible]	[illegible] 0.1335	0.198 ± 0.1063	0.198 ± 0.1063
	(UB) LOG LIKELYHOOD	0.035 ± 0.0368	0.0095 ± 0.0195	**0.0046 ± 0.003**	0.0429 ± 0.043	0.1102 [illegible] 0.06[illegible]	[illegible].1102 ± [illegible].0607	0.1117 ± 0.0714	0.1117 ± 0.0714
	(UB) PEARSON WEIGHTED	0.1139 ± 0.0439	0.0385 ± 0.0452	**0.0146 ± 0.0129**	0.0893 ± 0.0532	0.1955 ± 0.076[illegible]	[illegible].1991 ± 0.0854	0.1581 ± 0.0597	0.1581 ± 0.0597

MEMBERS SATISFACTION DISPERSION (lower is better)									DATASET: YELP
		MAGReS [One-Step; M19]	MAGReS [MCP Strict; M5]	MAGReS [MCP Relaxed; M17]	MAGReS [MCP Next; M11]	TRADGRec-PA [AVG; T1]	TRADGRec-PA [UL; T5]	TRADGRec-RA [AVG; T1]	TRADGRec-RA [UL; T5]
SIMILARITY METRIC	(IB) CITY BLOCK	0.1015 ± 0.0546	**0.0538 ± 0.0693**	0.094 ± 0.0507	0.09[illegible] 0.0512	0.0919 ± 0.0413	0.0905 ± 0.0388	0.1144 ± 0.0611	0.1144 ± 0.0611
	(IB) EUCLIDEAN WEIGHTED	0.1633 ± 0.0712	0.0579 ± 0.0271	**0.0554 [illegible] 0.0237**	0.0809 ± [illegible].0632	0.2907 ± 0.154	0.2875 ± 0.1747	0.197 ± 0.0835	0.197 ± 0.0835
	(IB) PEARSON WEIGHTED	0.1419 ± 0.0725	0.0242 ± 0.0229	**0.0222 ± [illegible]**	0.04[illegible] 0.0621	0.2048 ± 0.0799	0.215 ± 0.0809	0.1965 ± 0.0881	0.1965 ± 0.0881
	(IB) UNCENTERED COSINE WEIGHTED	0.1639 ± 0.072	**0.0579 ± 0.0257**	0.06[illegible] ± 0.025[illegible]	0.0806 ± 0.0579	0.288 ± 0.1557	0.2895 ± 0.1731	0.1969 ± 0.0835	0.1969 ± 0.0835
	(UB) LOG LIKELYHOOD	0.1436 ± 0.0718	**0.0004 ± 0.0026**	[illegible]0051 ± [illegible]	0.0133 ± 0.0325	0.2898 ± 0.1105	0.2898 ± 0.1105	0.2113 ± 0.077	0.2003 ± 0.0791
	(UB) PEARSON WEIGHTED	0.116 ± 0.0868	**0.0128 ± 0.0[illegible]**	0.[illegible]44 ± 0.0[illegible]8	0.0321 ± 0.0852	0.1544 ± 0.0864	0.1339 ± 0.0791	0.1431 ± 0.0817	0.1431 ± 0.0817

Figure 7: Average member satisfaction dispersion (MSD) per similarity metric and approach

FAIRNESS (higher is better)									DATASET: ML_LATEST_SMALL
		MAGReS [One-Step; M19]	MAGReS [MCP Strict; M5]	MAGReS [MCP Relaxed; M17]	MAGReS [MCP Next; M11]	TRADGRec-PA [AVG; T1]	TRADGRec-PA [UL; T5]	TRADGRec-RA [AVG; T1]	TRADGRec-RA [UL; T5]
SIMILARITY METRIC	(IB) CITY BLOCK	[illegible].6596 ± 0.193[illegible]	0.5864 ± [illegible].046	0.6912 ± 0.1971	0.6775 ± 0.1982	0.6609 ± 0.2166	0.6621 ± 0.22	**0.6857 ± 0.1923**	**0.6857 ± 0.1923**
	(IB) EUCLIDEAN WEIGHTED	0.8[illegible] 0.1332	[illegible].8862 ± 0.2147	**0.9117 ± 0.1008**	0.8864 ± 0.1073	0.6357 ± 0.1724	0.6383 ± 0.1589	0.6922 ± 0.144	0.6922 ± 0.144
	(IB) PEARSON WEIGHTED	0.7204 ± [illegible].115	0.9738 ± 0.0419	**0.9853 ± 0.0237**	0.0519 ± 0.0418	0.5506 ± 0.1118	0.5556 ± 0.1202	0.6706 ± 0.0921	0.6706 ± 0.0921
	(IB) UNCENTERED COSINE WEIGHTED	0.797[illegible] 0.1347	0.9132 ± 0.1692	**0.9246 ± 0.1002**	0.8748 ± 0.1241	0.6237 ± 0.1808	0.6216 ± 0.169	0.6963 ± 0.1359	0.6963 ± 0.1359
	(UB) LOG LIKELYHOOD	0.9743 ± 0.0392	0.9982 ± 0.0094	**1 ± 0**	0.9669 ± 0.044	0.8788 ± 0.0564	0.8788 ± 0.0564	0.8987 ± 0.0723	0.8987 ± 0.0723
	(UB) PEARSON WEIGHTED	0.7486 ± 0.0963	0.982 ± 0.0392	**0.9996 ± 0.003**	0.9204 ± 0.0507	0.5624 ± 0.0943	0.5991 ± 0.1039	0.6814 ± 0.0869	0.6814 ± 0.0869

FAIRNESS (higher is better)									DATASET: YELP
		MAGReS [One-Step; M19]	MAGReS [MCP Strict; M5]	MAGReS [MCP Relaxed; M17]	MAGReS [MCP Next; M11]	TRADGRec-PA [AVG; T1]	TRADGRec-PA [UL; T5]	TRADGRec-RA [AVG; T1]	TRADGRec-RA [UL; T5]
SIMILARITY METRIC	(IB) CITY BLOCK	0.822 ± 0.2032	0.429 ± 0.4651	0.8449 ± 0.199	**0.8498 ± 0.1885**	0.8438 ± 0.1927	0.8404 ± 0.1957	0.8259 ± 0.2055	0.8259 ± 0.2055
	(IB) EUCLIDEAN WEIGHTED	0.6952 ± 0.1592	0.975 ± 0.0564	**0.9838 ± 0.0426**	0.9341 ± 0.0645	0.1879 ± 0.0988	0.2009 ± 0.1216	0.6092 ± 0.1229	0.6092 ± 0.1229
	(IB) PEARSON WEIGHTED	0.6771 ± 0.1092	0.9959 ± 0.01	**0.9959 ± 0.0148**	0.9666 ± 0.0613	0.2854 ± 0.105	0.2928 ± 0.1264	0.5796 ± 0.0968	0.5796 ± 0.0968
	(IB) UNCENTERED COSINE WEIGHTED	0.697 ± 0.1461	0.9816 ± 0.0446	**0.9833 ± 0.041**	0.9311 ± 0.0671	0.637 ± 0.1906	0.2036 ± 0.1175	0.6088 ± 0.123	0.6088 ± 0.123
	(UB) LOG LIKELYHOOD	0.736 ± 0.1189	**1 ± 0**	**1 ± 0**	0.9904 ± 0.0269	0.3881 ± 0.0992	0.3881 ± 0.0992	0.6669 ± 0.1097	0.6744 ± 0.0861
	(UB) PEARSON WEIGHTED	0.8211 ± 0.0944	0.9926 ± 0.0497	**0.9926 ± 0.0497**	0.9758 ± 0.0717	0.7272 ± 0.0921	0.7322 ± 0.1151	0.7771 ± 0.0824	0.7771 ± 0.0824

Figure 8: Average fairness per similarity metric and approach

M19] versus *MAGReS [MCP Strict; M5]* or *MAGReS [MCP Relaxed; M17]*) versus *TRADGRec-PA [AVG; T1]*). To do this, we first used a two-sided test to determine if both samples were significantly different from one another and then, if the samples were different, we used two one-sided tests to determine which of the samples was greater or less than the other. For the experiments in Figure 6 we wanted to test whether one approach, A (e.g., *MAGReS [MCP Strict; M5]*), was significantly better than the other B (e.g. *TRADGRec-PA [AVG; T1]*) in terms of GS, so the null hypothesis was "the sample of the approach A is not greater than the sample of the approach B". For the experiments reported in Figure 7 we wanted to test that one approach A was significantly better than the other B in terms of MSD, and therefore the null hypothesis was "the sample of the approach A is not worse than the sample of the approach B". Finally, for the experiments reported in Figure 8 we wanted to test whether one approach A was significantly better than other approach B in terms of fairness, so the null hypothesis was "the sample of the approach A is not greater than the sample of the approach B". The statistical analysis confirmed our observations as, in each test, the null hypothesis was rejected at a significance level of 95% ($\alpha = 0.05$). Thus, we confirmed that the recommendations generated by MAGReS, and particularly those generated by *MAGReS [MCP Relaxed; M17]* outperformed those of *TRADGRec-PA* and *TRADGRec-RA*.

4.5.2. Already Rated Punishment Strategy

The *ARP* strategy was created to model how the user feels about receiving a recommendation with an item she has already rated. Each variant of the *ARP* strategy affects the way the proposals are treated by the agents, both when they decide about accepting (or rejecting) a proposal and about their next proposal, and this might ultimately reduce the amount of cases in which an item already rated is recommended to the group. We tested all the strategies using both user and item-based SURs, and each of them with different similarity metrics. The parameterization used for this experiment is detailed in Appendix A.4.2. At the moment, all the agents use the same ARP Strategy, but we plan to allow users to choose their own strategy in their agents in the near future.

Figure 9 shows, per *ARP* strategy, the average percentage of items recommended (considering the 45 groups per dataset) that were already rated by at least one of the group members. As it can be seen, the tests revealed that when using ML_LATEST_SMALL dataset, *Taboo* is the best-perfoming strategy at the task of reducing the amount of already-rated items being recommended. This implies that the desires of the users who chose the *Taboo* as the ARP Strategy for their agents are being properly represented and taken into account during the recommendation process. The tests with the YELP dataset confirmed what we observed in the tests with the other dataset, but also showed that when a user-based similarity metric was used the overlap was close to zero even when using the other ARP strategies. This may be explained by the amount of items in the dataset (as the POI dataset has almost 10 times more items than the movies dataset) and the amount of items rated by each user (a minimum of 9 items per user in YELP dataset, and a minimum of 20

OVERLAP (%) (lower is better) — SIMILARITY METRIC	DATASET: ML_LATEST_SM[illegible] EASY-GOING	FLEXIBLE [f=0.75]	FLEXIBLE PLUS [f=0.75. ms=0.6]	MIN-SATISFACTION [ms=0.6]	T[illegible]
(IB) CITY BLOCK	31.78% ± 32.77%	11.49% ± 25.65%	24.39% ± 31.05%	31.89% ± 32.83%	**4.28% ± 17.34[illegible]**
(IB) EUCLIDEAN WEIGHTED	6.72% ± 16.02%	0.84% ± 6.24%	1.14% ± 6.97%	6.94% ± 16.54%	[illegible]
(IB) PEARSON WEIGHTED	2.06% ± 7.3%	0.06% ± 0.75%	0.22% ± 2.98%	2[illegible] 6[illegible]%	**0% ± [illegible]**
(IB) UNCENTERED COSINE WEIGHTED	4.63% ± 12.7%	0.86% ± 6.04%	2.22% ± 12.58%	3.[illegible]% ± 11.[illegible]%	**0% ± 0%**
(UB) LOG LIKELYHOOD	2.72% ± 8.11%	**0% ± 0%**	**0% ± 0%**	[illegible].72% ± [illegible].11%	**[illegible] ± 0%**
(UB) PEARSON WEIGHTED	8.06% ± 12.1%	0.11% ± 1.49%	0.44% ± 2.32%	8.5[illegible] [illegible]36%	**0% ± 0%**

OVERLAP (%) (lower is better) — SIMILARITY METRIC	[illegible]ATASET: YELP EASY-GOING	FLEXIBLE [f=0.75]	FLEXIBLE [illegible] [f=0.75. ms=0.6]	M[illegible] [illegible]TISFACTION [illegible]=0.6]	TABOO
(IB) CITY BLOCK	14.96% ± 31.16%	1.78% ± 10.07%	10.19% ± [illegible]3%	1[illegible]6% ± [illegible].16%	**0% ± 0%**
(IB) EUCLIDEAN WEIGHTED	2.08% ± 6.65%	0.09% ± 1.24%	0.71% ± [illegible].89%	1.7% ± 6.07%	**0% ± 0%**
(IB) PEARSON WEIGHTED	2.28% ± 7.76%	0.06% ± [illegible]	0.41[illegible]	2.73% ± 8.19%	**0% ± 0%**
(IB) UNCENTERED COSINE WEIGHTED	2.43% ± 6.35%	0.16[illegible] 1.55%	0.97% ± 4.45%	2.31% ± 6.29%	**0% ± 0%**
(UB) LOG LIKELYHOOD	0.78% ± 4.66%	**[illegible]**	**[illegible]% ± 0%**	0.78% ± 4.66%	**0% ± 0%**
(UB) PEARSON WEIGHTED	**0% ± 0%**	**0[illegible]**	**0% ± 0%**	**0% ± 0%**	**0% ± 0%**

Figure 9: Average overlap percentage per similarity metric and ARP strategy.

in ML_LATEST_SMALL), which reduces the probability of occurrence of the overlaps.

In Figure 9 we also observed that using *Min-Satisfaction* with the min-satisfaction parameter set to 0.6 (equivalent to 3 of 5 stars) is almost the same as using *Easy-Going* in terms of both group satisfaction and percentage of overlap. This means that even when the agents are not penalizing proposals with less than 3 stars, since they are using the *Easy-Going* strategy that constraint is being maintained implicitly. Finally, we observed that *Flexible Plus* seems to be a good point in between an aggressive strategy like *Taboo* and a relaxed strategy like *Easy-Going*, helping to discard items (e.g., movies) the user will not like but still recommending those she will be willing to consume again (e.g., movies the user will be willing to watch again). The use of the *ARP* strategies has a side-effect on the group satisfaction value: the more restrictive the strategy is, the lower the group satisfaction is. The reason behind this observation is that when users (and therefore, their agents) do not "complain" (by penalizing the utility reported) about receiving proposals with items they have already rated (i.e. when the *Easy-Going* strategy is used), it is more likely that the recommendations contain items that received high ratings by the group members, which then increases the group satisfaction value (see Figure 10). This way, *Easy-Going* produces the best recommendations in terms of group satisfaction, but the worst one in terms of the overlap , and *Taboo* produces recommendations with less overlap but also with an slightly lower value of group satisfaction.

With regard to the MSD and fairness (see Figures 11 and 12) of the recommendations, the results vary depending on the dataset. For the dataset

GROUP SATISFACTION (higher is better)	DATASET: ML_LATEST_SM[illegible]				
SIMILARITY METRIC	EASY-GOING	FLEXIBLE [f=0.75]	FLEXIBLE PLUS [f=0.75. ms=0.6]	MIN-SATISFACTION [ms=0.6]	T[illegible]
(IB) CITY BLOCK	0.7704 ± 0.0563	0.7006 ± 0.1965	**0.7736 ± 0.0604**	0.7704 ± 0.0563	0.6297 ± 0.2547
(IB) EUCLIDEAN WEIGHTED	**0.8463 ± 0.0452**	0.8266 ± 0.1334	0.8264 ± 0.1334	0.8458 ± 0.0456	[illegible] ± 0.1[illegible]
(IB) PEARSON WEIGHTED	0.9504 ± 0.0434	0.9495 ± 0.0433	0.9501 ± 0.043	0.94[illegible] ± 0.[illegible]23	**0.9519 ± 0.0[illegible]7**
(IB) UNCENTERED COSINE WEIGHTED	**0.8441 ± 0.0457**	0.8261 ± 0.1335	0.8448 ± 0.0442	0.8[illegible]36 ± 0.[illegible]	0.8256 ± 0.133[illegible]
(UB) LOG LIKELYHOOD	0.9649 ± 0.0175	0.9672 ± 0.0113	**0.9686 ± 0.0128**	0.9649 ± 0.0175	[illegible]2 ± 0.0113
(UB) PEARSON WEIGHTED	**0.9274 ± 0.0235**	0.9214 ± 0.0358	0.9223 ± 0.0351	0.9[illegible] ± [illegible]236	0.9206 ± 0.0378

GROUP SATISFACTION (higher is better)	[illegible]ATASET: YELP				
SIMILARITY METRIC	EASY-GOING	FLEXIBLE [f=0.75]	FLEXIBLE [illegible] [f=0.75. ms=0.6]	MI[illegible] [illegible]TISFACTION [illegible]=0.6]	TABOO
(IB) CITY BLOCK	**0.6233 ± 0.3632**	0.4284 ± 0.4093	0.5494 ± [illegible]967	**0[illegible]233 ± [illegible]3632**	0.3902 ± 0.4078
(IB) EUCLIDEAN WEIGHTED	0.9084 ± 0.0475	0.906 ± 0.0484	0.9075 ± [illegible]048	**0.9088 ± 0.048**	0.9057 ± 0.0488
(IB) PEARSON WEIGHTED	**0.9799 ± 0.0234**	0.9778 ± [illegible]	0.978[illegible] ± [illegible]	0.9797 ± 0.0233	0.9773 ± 0.0249
(IB) UNCENTERED COSINE WEIGHTED	**0.9098 ± 0.0473**	0.906[illegible] ± 0.0473	0.9086 ± 0.0464	0.9097 ± 0.0465	0.9064 ± 0.0475
(UB) LOG LIKELYHOOD	0.9988 ± 0.0067	[illegible]988 ± 0.00[illegible]	0.[illegible]988 ± 0.0067	**0.9989 ± 0.0061**	**0.9989 ± 0.0064**
(UB) PEARSON WEIGHTED	**0.9926 ± 0.0497**	**[illegible]99[illegible] ± [illegible]97**	**0.9926 ± 0.0497**	**0.9926 ± 0.0497**	**0.9926 ± 0.0497**

Figure 10: Average group satisfaction per similarity metric and ARP strategy.

ML_LATEST_SMALL, we can say that *Taboo* is the best ARP strategy with regard to MSD but *Min-Satisfaction* (with the *ms* parameter set to 0.6) is the best with regard to fairness. For YELP, *Taboo* is the best with regard to the fairness but there is not a clear winner regarding to the MSD. Overall, and independently of the dataset, we would choose *Taboo* as a default strategy (until the user chooses the strategy that better suits her personality) as, in most cases, it ensures the minimization of the overlap.

All the results were validated through statistical tests. We first run the Shappiro-Wilk test on the samples to determine whether they followed or not the normal distribution. Given that some of the samples did not follow that distribution a non-parametric test was used. We then proceeded like we did previously (see Section 4.5.1): by performing (for each dataset and for each similarity metric) a pair-wise comparisons among the 5 samples (one for each *ARP* variant) using the Wilcoxon Signed Ranks test. For Figures 9 and 11 the null hypothesis was "the sample of the variant A is not worse than the sample of the variant B", as we wanted to test that one *ARP* variant A (e.g., *Taboo*) was significantly better than the other B (e.g., *Easy-Going*) in terms of the amount of overlap (Figure 9)/MSD (Figure 11). In the case of the Figures 10 and 12 the null hypothesis was "the sample of the variant A is not greater than the sample of the variant B". The test results confirmed our observations as, in each test, the null hypothesis was rejected at a significance level of 95%. Thus, we confirmed that the variants of the ARP proposed in this article helped to reduce the overlap in the recommendations, but at the cost of causing a minor loss in their quality.

MEMBERS SATISFACTION DISPERSION (lower is better) — SIMILARITY METRIC	DATASET: ML_LATEST_SMALL EASY-GOING	FLEXIBLE [f=0.75]	FLEXIBLE PLUS [f=0.75. ms=0.6]	MIN-SATISFACTION [ms=0.6]	TABOO
(IB) CITY BLOCK	0.0902 ± 0.0472	0.0819 ± 0.0559	0.0966 ± 0.0492	0.0917 ± 0.0472	**[illegible]**
(IB) EUCLIDEAN WEIGHTED	0.0749 ± 0.0502	0.0751 ± 0.0498	0.0756 ± 0.0497	[illegible]	**0.0745 ± [illegible]**
(IB) PEARSON WEIGHTED	0.046 ± 0.0514	0.0458 ± 0.0506	0.0449 ± 0.0502	[illegible]	**[illegible]**
(IB) UNCENTERED COSINE WEIGHTED	0.0802 ± 0.0444	**0.0777 ± 0.0485**	0.0811 ± 0.0468	[illegible]	[illegible]
(UB) LOG LIKELYHOOD	0.0124 ± 0.0207	0.0095 ± 0.0195	**0.0093 ± 0.0196**	[illegible]	0.0095 ± 0.0195
(UB) PEARSON WEIGHTED	**0.0197 ± 0.0089**	0.039 ± 0.0457	0.0382 ± [illegible]	0.0337 ± 0.0247	0.0385 ± 0.0452

MEMBERS SATISFACTION DISPERSION (lower is better) — SIMILARITY METRIC	DATASET: YELP EASY-GOING	FLEXIBLE [f=0.75]	FLEXIBLE PLUS [f=0.75. ms=0.6]	MIN-SATISFACTION [ms=0.6]	TABOO
(IB) CITY BLOCK	0.09 ± 0.0741	0.0613 ± 0.0727	[illegible] ± 0.0765	0.09 ± 0.0741	**0.0538 ± 0.0693**
(IB) EUCLIDEAN WEIGHTED	0.0572 ± 0.0259	0.0576 ± [illegible]	**[illegible]**	0.0573 ± 0.0264	0.0579 ± 0.0271
(IB) PEARSON WEIGHTED	**0.0221 ± 0.0224**	[illegible] ± 0.0226	0.0235 ± 0.0223	**0.0221 ± 0.0226**	0.0242 ± 0.0229
(IB) UNCENTERED COSINE WEIGHTED	**0.0561 ± 0.0269**	0.0577 ± [illegible]	[illegible] ± 0.0264	0.0563 ± 0.0266	0.0579 ± 0.0257
(UB) LOG LIKELYHOOD	0.0006 ± 0.0028	**[illegible]**	**0.0003 ± 0.0017**	0.0006 ± 0.0028	0.0004 ± 0.0026
(UB) PEARSON WEIGHTED	**[illegible]**	**[illegible]**	**0.0128 ± 0.0861**	**0.0128 ± 0.0861**	**0.0128 ± 0.0861**

Figure 11: Average MSD per similarity metric depending on the ARP strategy used by all the group members

FAIRNESS (higher is better) — SIMILARITY METRIC	DATASET: ML_LATEST_SMALL EASY-GOING	FLEXIBLE [f=0.75]	FLEXIBLE PLUS [f=0.75. ms=0.6]	MIN-SATISFACTION [ms=0.6]	TABOO
(IB) CITY BLOCK	0.761 ± 0.1818	0.6645 ± 0.2612	0.7596 ± 0.1877	**0.7646 ± 0.1816**	0.5864 ± 0.3046
(IB) EUCLIDEAN WEIGHTED	0.9 ± 0.1689	0.9032 ± 0.1771	0.9023 ± 0.1766	**0.9196 ± 0.1021**	0.8862 ± 0.2147
(IB) PEARSON WEIGHTED	0.9698 ± 0.0515	0.9679 ± 0.0507	0.9705 ± 0.0512	**0.9748 ± 0.0472**	0.9738 ± 0.0419
(IB) UNCENTERED COSINE WEIGHTED	0.9295 ± 0.0977	0.9126 ± 0.175	**0.9377 ± 0.0965**	0.9269 ± 0.1011	0.9132 ± 0.1692
(UB) LOG LIKELYHOOD	**0.9987 ± 0.0089**	0.9982 ± 0.0094	0.9982 ± 0.0094	**0.9987 ± 0.0089**	0.9982 ± 0.0094
(UB) PEARSON WEIGHTED	**0.9996 ± 0.003**	0.9816 ± 0.0367	0.9816 ± 0.0367	0.9904 ± 0.017	0.982 ± 0.0392

FAIRNESS (higher is better) — SIMILARITY METRIC	DATASET: YELP EASY-GOING	FLEXIBLE [f=0.75]	FLEXIBLE PLUS [f=0.75. ms=0.6]	MIN-SATISFACTION [ms=0.6]	TABOO
(IB) CITY BLOCK	**0.6497 ± 0.4086**	0.4657 ± 0.461	0.57 ± 0.4362	**0.6497 ± 0.4086**	0.429 ± 0.4651
(IB) EUCLIDEAN WEIGHTED	**0.9771 ± 0.0539**	0.9758 ± 0.054	0.9768 ± 0.054	0.9761 ± 0.0606	0.975 ± 0.0564
(IB) PEARSON WEIGHTED	0.9953 ± 0.0126	0.9954 ± 0.0105	0.9945 ± 0.0115	0.9951 ± 0.011	**0.9959 ± 0.01**
(IB) UNCENTERED COSINE WEIGHTED	**0.9844 ± 0.0428**	0.9822 ± 0.0447	0.9829 ± 0.0447	0.9841 ± 0.0436	0.9816 ± 0.0446
(UB) LOG LIKELYHOOD	**1 ± 0**	**1 ± 0**	**1 ± 0**	**1 ± 0**	**1 ± 0**
(UB) PEARSON WEIGHTED	**0.9926 ± 0.0497**	**0.9926 ± 0.0497**	**0.9926 ± 0.0497**	**0.9926 ± 0.0497**	**0.9926 ± 0.0497**

Figure 12: Average fairness per similarity metric depending on the ARP strategy used by all the group members

AMOUNT OF CONCESSIONS (per recommendation. lower is better)	DATASET: ML_LATEST_SMALL				
SIMILARITY METRIC	STRICT	RELAXED [rp=0.025]	RELAXED [rp=0.05]	RELAXED [rp=0.1]	NEXT
(IB) CITY BLOCK	38.9944 ± 53.627	28.7667 ± 51.0631	10.7444 ± 26.6792	**2.5944 ± 7.0336**	[illegible]2.9667 ± 53.435
(IB) EUCLIDEAN WEIGHTED	60.5222 ± 35.8288	59.1222 ± 34.727	56.5167 ± 34.4968	**43.0278 ± 34.0513**	[illegible]
(IB) PEARSON WEIGHTED	55.7222 ± 54.5227	57.0389 ± 55.8777	53.8667 ± 56.2164	**[illegible]**	46.8444 ± [illegible]
(IB) UNCENTERED COSINE WEIGHTED	49.7833 ± 27.7109	47.2556 ± 26.6756	44.7389 ± 25.2855	**[illegible]**	[illegible]
(UB) LOG LIKELYHOOD	13.6611 ± 7.9003	12.1333 ± 6.5429	9.6444 ± 6.4102	**[illegible]**	[illegible]
(UB) PEARSON WEIGHTED	50.7333 ± 22.564	42.7556 ± 17.21	[illegible]	**[illegible]**	25.6778 ± 14.4252

AMOUNT OF CONCESSIONS (lower is better)	DATASET: YELP				
SIMILARITY METRIC	STRICT	RELAXED [rp=0.025]	RELAXED [rp=0.05]	RELAXED [rp=0.1]	NEXT
(IB) CITY BLOCK	37.25 ± 90.5903	68.2944 ± 143.4417	52.6556 ± [illegible]	**[illegible]**	112.6167 ± 207.5576
(IB) EUCLIDEAN WEIGHTED	107.9944 ± 61.7056	[illegible]	[illegible]	**[illegible] ± 74.028**	99.6389 ± 70.2895
(IB) PEARSON WEIGHTED	46.9 ± 59.3681	42.0444 ± 59.4557	[illegible]	**36.2611 ± 62.5769**	41.3611 ± 62.4437
(IB) UNCENTERED COSINE WEIGHTED	98.8 ± 58.394	[illegible]	101.5667 ± 65.6866	**93.2667 ± 64.3929**	92.0389 ± 68.1616
(UB) LOG LIKELYHOOD	9.8 ± 17.4892	[illegible]	[illegible] ± 14.4676	**7.0111 ± 12.6381**	8.8444 ± 15.705
(UB) PEARSON WEIGHTED	2.9889 ± 5.0539	[illegible]	2.7556 ± 4.5973	**2.6111 ± 4.1714**	2.7111 ± 4.797

Figure 13: Average amount of concessions to generate a recommendation of $k = 10$ items depending on the PrA strategy

4.5.3. Proposals Acceptance strategy

In addition to the (existing) *Strict* variant, we proposed two more strategies, namely: *Relaxed* and *Next*. Although the usage of the *PrA* strategy is merely dedicated to model acceptance criteria for user's proposals, we will compare the different variants of the strategy according to how they affect recommendation process and results. By definition, the *Relaxed* and *Next* variants of the *PrA* strategy cause, in some way, the agents to be more "flexible" when deciding whether to accept a proposal. This situation has the following consequences:

1. It helps the agents to reach agreements faster and therefore produce recommendations faster. Given that the agents are "more relaxed" when assessing proposals and deciding whether to accept them , less concessions are needed to reach an agreement and the number of rounds of negotiation consequently decreases. As it can be seen in Figure 13 there are some exceptions to the rule, for example for ML_LATEST_SMALL when the *(IB) City Block* similarity metric is used, and for YELP when the *(IB) Euclidean Weighted* similarity metric was used. In both cases the increase in the number of concessions when using the strategies *Relaxed* and *Next* can be explained by the difference in the GS of the recommendations produced when those *PrA* strategies were used.
2. It might increase the group satisfaction for certain configurations of groups and strategies (see Figure 14). This situation is highly dependent on the strategy followed by the agents to select their initial proposal and the *Concession* strategies. For the experiments we used the *Egocentric* [13] (the

initial proposal is the one that retains the highest utility value) and *Desires Distance* strategies respectively. In this context, each agent´s initial proposal is its best one in terms of utility (and therefore user satisfaction) and every time the agent has to concede it makes a new proposal with lower utility than its current one. Then, the more concessions we have, the lower the user satisfaction will be. Moreover, each concession lowers the "agent's requirements" for accepting proposals and increases the probability of reaching an agreement. This way, the more concessions the agent makes, the lower its utility becomes, and therefore the lower the satisfaction of the corresponding user will be.

GROUP SATISFACTION (higher is better)	DATASET: ML_LATEST_SMALL				
SIMILARITY METRIC	STRICT	RELAXED [rp=0.02[illegible]	[illegible] [rp=0.[illegible]	[illegible]LAXED [rp=0.1]	NEXT
(IB) CITY BLOCK	0.6297 ± 0.2547	0.697 ± 0.1592	[illegible]284 ± 0.[illegible]	0.7282 ± 0.0465	**0.7303 ± 0.0499**
(IB) EUCLIDEAN WEIGHTED	0.8103 ± 0.182	[illegible] ± 0.1[illegible]	0.8465 ± 0.0443	**0.8472 ± 0.044**	0.8207 ± 0.0689
(IB) PEARSON WEIGHTED	0.9519 ± 0.0387	0.9553 ± 0.0369	**[illegible].9566 ± [illegible]0357**	0.9545 ± 0.0318	0.9403 ± 0.0529
(IB) UNCENTERED COSINE WEIGHTED	0.8256 ± 0.1336	0.844[illegible] ± 0.04[illegible]	[illegible].8453 ± 0.0452	**0.8464 ± 0.0452**	0.8141 ± 0.0748
(UB) LOG LIKELYHOOD	0.9[illegible]72 ± [illegible]	**[illegible]88 ± 0.[illegible]3**	0.9684 ± 0.0074	0.9681 ± 0.0073	0.9415 ± 0.0352
(UB) PEARSON WEIGHTED	0.920[illegible] ± 0.0378	[illegible].9354 ± [illegible]88	0.9354 ± 0.0194	**0.9371 ± 0.019**	0.8753 ± 0.0383

GROUP SATISFACTION (higher is better)	DATASET: YELP				
SIMILARITY METRIC	S[illegible]T	RELAXED [rp=0.025]	RELAXED [rp=0.05]	RELAXED [rp=0.1]	NEXT
(IB) CITY BLOCK	0.3902 ± 0.4078	0.7072 ± 0.2615	0.7572 ± 0.1791	0.7885 ± 0.0719	**0.7895 ± 0.0709**
(IB) EUCLIDEAN W[illegible]GHTED	0.9057 ± [illegible].0488	0.9076 ± 0.0471	0.9108 ± 0.0464	**0.9116 ± 0.0422**	0.8769 ± 0.0608
(IB) PEARSON WE[illegible]TED	[illegible]9773 ± [illegible].0249	0.9786 ± 0.0245	**0.9787 ± 0.0246**	0.9765 ± 0.0219	0.964 ± 0.0531
(IB) UNC[illegible]TERED COSINE [illegible]EIGHTED	0.9064 ± 0.0475	0.9071 ± 0.0466	0.9089 ± 0.0449	**0.9108 ± 0.0436**	0.8755 ± 0.0602
(U[illegible]OG LIK[illegible]HOOD	**0.9989 ± 0.0064**	0.9987 ± 0.0062	0.9976 ± 0.0064	0.9943 ± 0.0075	0.9896 ± 0.0293
[illegible]B) PEARSON W[illegible]D	**0.9926 ± 0.0497**	0.9925 ± 0.0497	0.9919 ± 0.0496	0.9909 ± 0.0495	0.9757 ± 0.069

Figure 14: Average group satisfaction (GS) depending on the PrA strategy

3. It might increase the amount of effective recommendations produced by the recommender to each group, specially when the similarity metric is an item based one (see Figure 16). This is more noticeable when all the agents use a "more relaxed" *PrA* strategy (like *Next* or *Relaxed*) and the similarity metric is an item-based one. In fact, as expected, the less strict the agent is when determining whether to accept or reject proposals, the higher the amount of effective recommendations is. The explanation for this is simple: if the agents are more prone to accepting proposals, more negotiations will end with an *agreement* and therefore more items will be recommended to the group.

The parameterization used for this experiment is specified in Appendix A.4.3. We have tested the strategies with different similarity metrics and *ARP* strategies, while keeping the *Initial Proposal* and *Concession* strategies fixed (to *Egocentric* and *Desires Distance* respectively). At the moment, all the agents use

MEMBERS SATISFACTION DISPERSION (lower is better)	DATASET: ML_LATEST_SMALL				
SIMILARITY METRIC	STRICT	RELAXED [rp=0.025]	RELAXED [rp=0.05]	RELAXED [rp=0.1]	NEXT
(IB) CITY BLOCK	**0.069 ± 0.0503**	0.0719 ± 0.0435	0.075 ± 0.0412	0.075 ± 0.0406	[illegible]
(IB) EUCLIDEAN WEIGHTED	**0.0745 ± 0.0517**	0.0782 ± 0.05	0.0835 ± 0.0502	[illegible]	0.112 ± [illegible]
(IB) PEARSON WEIGHTED	0.0395 ± 0.0306	0.0372 ± 0.0311	0.0375 ± 0.0323	[illegible]	[illegible]
(IB) UNCENTERED COSINE WEIGHTED	**0.0803 ± 0.0487**	0.081 ± 0.0489	0.0821 ± 0.0505	0.0847 ± 0.0522	[illegible] ± 0.0935
(UB) LOG LIKELYHOOD	0.0095 ± 0.0195	0.0054 ± 0.0037	0.0057 ± 0.0035	[illegible]	0.0429 ± 0.043
(UB) PEARSON WEIGHTED	0.0385 ± 0.0452	0.0174 ± 0.01	0.0174 ± [illegible]	**0.0146 ± [illegible]**	0.0893 ± 0.0532

MEMBERS SATISFACTION DISPERSION (lower is better)	DATASET: YELP				
SIMILARITY METRIC	STRICT	RELAXED [rp=0.025]	RELAXED [rp=0.05]	RELAXED [rp=0.1]	NEXT
(IB) CITY BLOCK	**0.0538 ± 0.0693**	0.0863 ± 0.0546	[illegible] ± 0.0489	0.094 ± 0.0507	0.0954 ± 0.0512
(IB) EUCLIDEAN WEIGHTED	0.0579 ± 0.0271	0.056 ± [illegible]	[illegible]	**0.0554 ± 0.0237**	0.0809 ± 0.0632
(IB) PEARSON WEIGHTED	0.0242 ± 0.0229	[illegible] ± 0.0247	0.0237 ± 0.024	**0.0222 ± 0.0185**	0.0404 ± 0.0621
(IB) UNCENTERED COSINE WEIGHTED	0.0579 ± 0.0257	**0.0572 ± [illegible]**	[illegible] ± 0.0279	0.0609 ± 0.0251	0.0806 ± 0.0579
(UB) LOG LIKELYHOOD	**0.0004 ± 0.0026**	[illegible]	0.002 ± 0.0031	0.0051 ± 0.0054	0.0133 ± 0.0325
(UB) PEARSON WEIGHTED	[illegible]	[illegible]	0.0137 ± 0.0859	0.0144 ± 0.0858	0.0321 ± 0.0852

Figure 15: Average member satisfa[illegible] disp[illegible]rsion (MSD) depending on the Proposals Acceptance PrA strategy used by all the g[illegible]up members

AMO[illegible]NT OF RECOM[illegible]ENDAT[illegible]NS (hi[illegible]er is bett[illegible])	DATASET: ML_LATEST_SMALL				
SIMILA[illegible]Y METRIC	STRICT	RELAXED [rp=0.025]	RELAXED [rp=0.05]	RELAXED [rp=0.1]	NEXT
(IB) [illegible]LOCK	7.0667 ± 3.84	9.4222 ± 2.1584	9.9111 ± 0.5963	**10 ± 0**	**10 ± 0**
[illegible]B) EUCLIDEAN WEIG[illegible]TED	9.3333 ± 2.1847	9.6 ± 1.8878	9.7111 ± 1.1604	9.7778 ± 1.0636	**10 ± 0**
(IB) PEA[illegible] WEIGHTED	**10 ± 0**	**10 ± 0**	**10 ± 0**	**10 ± 0**	**10 ± 0**
(IB) UNCENTERED COSINE WEIGHTED	9.4222 ± 2.0393	9.5111 ± 1.9024	9.7111 ± 1.2725	9.8889 ± 0.7454	**10 ± 0**
(UB) LOG LIKELYHOOD	**10 ± 0**	**10 ± 0**	**10 ± 0**	**10 ± 0**	**10 ± 0**
[illegible]B) PEARSON WEIGHTED	**10 ± 0**	**10 ± 0**	**10 ± 0**	**10 ± 0**	**10 ± 0**

AMOUNT OF [illegible]ECOMMENDATIONS (higher is better)	DATASET: YELP				
SIMILARITY METRIC	STRICT	RELAXED [rp=0.025]	RELAXED [rp=0.05]	RELAXED [rp=0.1]	NEXT
(IB) CITY BLOCK	2.7778 ± 3.6549	7.6889 ± 3.4826	9.0444 ± 2.5312	9.7778 ± 1.0848	**10 ± 0**
(IB) EUCLIDEAN WEIGHTED	8.1778 ± 2.0702	8.7556 ± 1.5249	9.2889 ± 1.1406	9.8667 ± 0.4573	**10 ± 0**
(IB) PEARSON WEIGHTED	9.6667 ± 1.3314	9.6 ± 1.3551	9.6667 ± 1.3314	9.7111 ± 1.325	**10 ± 0**
(IB) UNCENTERED COSINE WEIGHTED	8.1111 ± 2.1237	8.6222 ± 1.6691	9.1333 ± 1.4397	9.7556 ± 0.529	**10 ± 0**
(UB) LOG LIKELYHOOD	9.9556 ± 0.2084	**10 ± 0**	**10 ± 0**	**10 ± 0**	**10 ± 0**
(UB) PEARSON WEIGHTED	**10 ± 0**	**10 ± 0**	**10 ± 0**	**10 ± 0**	**10 ± 0**

Figure 16: Average amount of effective recommendations depending on the PrA strategy

the same PrA Strategy, but we plan to allow users to choose their own strategy in their agents in the near future. Figures 13, 14, 15, 16 and 17 show the results for the test with the *Taboo ARP* strategy and the most relevant similarity metrics.

Figure 13 shows that, as expected, the total amount of concessions decreases drastically when using the *Relaxed [rp=0.1]* variant with respect to the *Strict* variant (which was the one used by PUMAS in [12]), and the rest of the variants are in between those mentioned.

In Figure 14 we see that group satisfaction for the *Relaxed* and *Next* variants did not increase as much as we expected. The results of the experiments with ML_LATEST_SMALL show the GS increased for all the similarity metrics. However, the experiments performed with YELP show that the GS only improved when an IB similarity metric was used. A possible explanation is that the *Relaxed* and *Next* variants of the *PrA* strategy assume that, in some cases, it might be preferable to accept a proposal x (even if it is not exactly better than the agent current proposal but it is close enough), rather than risking "to let" the agent concede (which will cause an utility loss perhaps higher than the one incurred by accepting proposal x). As a result, the agents accept more proposals, the agreement is reached faster and less concessions are made. Although this might be positive in some cases, there is a cost to pay: even if a proposal x is rejected, the agent might never be forced to concede and, in such a case, accepting x was a bad choice as it generated a utility loss that rejecting x would not have caused.

In Figure 15 we see that, in some cases the MSD decreases as the relax level increases (when using the *Relaxed* variant) but this is not always true. This effects follows from what we explained previously about the possible increase in group satisfaction but only in those cases on which the utility of the initial proposal of the agents is very similar and/or the same. Additionally, we can observe that the *Next* variant negatively impacted on how uniformly the group members were satisfied (by increasing the MSD), and once again, the explanation is the same as the one given above: the assumption that *Next* makes might lead the agents to make sub-optimal decisions.

Figure 16 shows that, as expected, when all the agents use the *Strict* variant the amount of effective recommendations (i.e recommendations produced) is lower than when they use the *Relaxed* or *Next variants*.

In Figure 17 we can see that the fairness of the recommendations increases in most cases when the *Relaxed* and *Next* strategies are used. All the results were validated through statistical tests. We first run the Shappiro-Wilk test on the samples to determine whether they followed or not the normal distribution. Given that some of the samples did not follow that distribution a non-parametric test was used. We then proceeded like we did previously (see Section 4.5.1): by performing (for each dataset and for each similarity metric) a pair-wise comparisons among the 5 samples (one for each *PrA* variant) using the Wilcoxon Signed Ranks test:

- For Figures 13 and 15, the null hypothesis was defined as "the sample of

FAIRNESS (higher is better)	DATASET: ML_LATEST_SM[illegible]				
SIMILARITY METRIC	STRICT	RELAXED [rp=0.025]	RELAXED [rp=0.05]	RELAXED [rp=0.1]	NEXT
(IB) CITY BLOCK	0.5864 ± 0.3046	0.6619 ± 0.2454	**0.6925 ± 0.1978**	0.6912 ± 0.1971	[illegible].6775 ± 0.1982
(IB) EUCLIDEAN WEIGHTED	0.8862 ± 0.2147	0.9046 ± 0.1656	**0.9212 ± 0.0936**	0.9117 ± 0.1008	[illegible]
(IB) PEARSON WEIGHTED	0.9738 ± 0.0419	0.9788 ± 0.029	0.9791 ± 0.0294	**[illegible]**	0.9405 ± [illegible]
(IB) UNCENTERED COSINE WEIGHTED	0.9132 ± 0.1692	**0.9383 ± 0.0903**	0.9295 ± 0.0954	[illegible]	[illegible] 0.124
(UB) LOG LIKELYHOOD	0.9982 ± 0.0094	**1 ± 0**	**1 ± 0**	**1 ± 0**	[illegible]
(UB) PEARSON WEIGHTED	0.982 ± 0.0392	0.9973 ± 0.0088	0.9971 ± 0.0086	**[illegible]**	0.9204 ± 0.0507

FAIRNESS (higher is better)	DATASET: YELP				
SIMILARITY METRIC	STRICT	RELAXED [rp=0.025]	RELAXED [rp=0.05]	RELAXED [rp=0.1]	NEXT
(IB) CITY BLOCK	0.429 ± 0.4651	0.7807 ± 0.3257	0.8193 ± 0.2588	[illegible]	**0.8498 ± 0.1885**
(IB) EUCLIDEAN WEIGHTED	0.975 ± 0.0564	0.9826 ± 0.0405	[illegible]	**[illegible]9838 ± 0.0426**	0.9341 ± 0.0645
(IB) PEARSON WEIGHTED	0.9959 ± 0.01	0.9962 ± 0.0105	[illegible]	0.9959 ± 0.0148	**0.9666 ± 0.0613**
(IB) UNCENTERED COSINE WEIGHTED	0.9816 ± 0.0446	[illegible]	**0.9841 ± 0.0406**	0.9833 ± 0.041	0.9311 ± 0.0671
(UB) LOG LIKELYHOOD	**1 ± 0**	**1 ± [illegible]**	**1 ± 0**	**1 ± 0**	0.9904 ± 0.0269
(UB) PEARSON WEIGHTED	**0.9926 ± 0.0497**	**[illegible]**	**0.9926 ± 0.0497**	**0.9926 ± 0.0497**	0.9758 ± 0.0717

Figure 17: Average fairness per similarity metric depending on the PrA Strategy

the variant A is not worse than the sample of the variant B" as we wanted to test that one *PrA* strategy A (e.g., *Strict*) was significantly better than the other B (e.g., *Next*) in terms of amount of concessions made by the agents (Figure 13)/MSD (Figure 15).

- For Figures 14, 16 and 17, we tested that one *PrA* strategy A (e.g., *Relaxed [rp=0.05]*) was significantly better than the other B (e.g., *Relaxed [rp=0.025]*) regarding to the GS (Figure 14)/amount of effective recommendations (Figure 16)/fairness (Figure 17), so null hypothesis was "the sample of the variant A is not greater than the sample of the variant B".

All the previous tests confirmed our observations as, in each test, the null hypothesis was rejected at a significance level of 95% ($\alpha = 0.05$). Thus, we confirmed that *PrA* Relaxed was the best one with regard to reducing the amount of concessions required for the agents to reach an agreement, and also to increasing the items recommended, while keeping the quality of the recommendations.

4.5.4. Information privacy

In order to measure the amount of information revealed by *UserAgent's* of MAGReS during the negotiation process, the UFIL and PIL indicators were computed (see Section 4.3).

The parameterization used for this experiment is specified in Appendix A.4.4. As it can be seen in Figure 18, the amount of information related to the utility function that was leaked when using MAGReS is always lower than when using the traditional approaches (the preference aggregation one, TRADGRec-PA,

INFORMATION LEAK - UTILITY FUNCTION (%) (lower is better)		MAGReS [One-Step; M19]	MAGReS [MCP Strict; M5]	MAGReS [MCP Relaxed; M17]	MAGReS [MCP Next; M11]	TRADGRec-PA [AVG; T1]	TRADGR[illegible]PA [U[illegible]5]	TRAD[illegible]RA [AVG; T1]	TRADGRec-RA [UL; T5]
									[illegible]TASET: ML[illegible]TEST_SMALL
SIMILARITY METRIC	(IB) CITY BLOCK	**1.97% ± 2.62%**	48.97% ± 36.02%	4.32% ± 3.5%	53.97% ± 26.84%	100% ± 0%	[illegible]0% ± [illegible]%	1[illegible] ± 0%	100% ± 0%
	(IB) EUCLIDEAN WEIGHTED	**0.88% ± 0.36%**	64.63% ± 17.73%	51.43% ± 20.28%	63.11% ± 14.54%	100% ± 0%	100% [illegible] 0%	100% ± 0%	100% ± 0%
	(IB) PEARSON WEIGHTED	**0.85% ± 0.01%**	31.12% ± 15.35%	24.99% ± 13.98%	41.07% ± 9.71%	100% [illegible] 0%	100% [illegible] 0%	100% ± 0%	100% ± 0%
	(IB) UNCENTERED COSINE WEIGHTED	**0.88% ± 0.36%**	64.08% ± 17.4%	54.4% ± 19.54%	37.09% ± 11.36%	100% ± 0%	100% ± [illegible]	100% ± 0%	100% ± 0%
	(UB) LOG LIKELYHOOD	**0.84% ± 0.01%**	16.91% ± 5.05%	4.86% ± 3.09%	19.67% ± 7.34%	100% ± 0%	100% ± 0%	100% ± 0%	100% ± 0%
	(UB) PEARSON WEIGHTED	**0.85% ± 0.01%**	29.17% ± 9.7%	14.35% ± 6.49%	26.96% ± 3.19%	[illegible] ± 0%	[illegible]00% ± 0%	100% ± 0%	100% ± 0%

INFORMATION LEAK - UTILITY FUNCTION (%) (lower is better)		MAGReS [One-Step; M19]	MAGReS [MCP Strict; M5]	MAGReS [MCP Relaxed; M17]	MAGReS [M[illegible]xt; M11]	TRADGR[illegible]PA [AVG; [illegible]]	TRADGRec-PA [UL; T5]	TRADGRec-RA [AVG; T1]	TRADGRec-RA [UL; T5]
									DATASET: YELP
SIMILARITY METRIC	(IB) CITY BLOCK	**0.15% ± 0.2%**	7.99% ± 10.01%	1.87% ± 4.87%	[illegible]% ± 4.18[illegible]	[illegible]0% ± 0%	100% ± 0%	100% ± 0%	100% ± 0%
	(IB) EUCLIDEAN WEIGHTED	**0.1% ± 0%**	14.46% ± 6.09%	9.48% ± 4.57%	[illegible] 3.59%	100% ± 0%	100% ± 0%	100% ± 0%	100% ± 0%
	(IB) PEARSON WEIGHTED	**0.1% ± 0.01%**	4.04% ± 2.55%	2.68% ± 2.[illegible]2%	[illegible]% ± 1.8[illegible]	100% ± 0%	100% ± 0%	100% ± 0%	100% ± 0%
	(IB) UNCENTERED COSINE WEIGHTED	**0.1% ± 0%**	14.57% ± 6.39%	10.[illegible]% ± 4.7[illegible]	[illegible] ± 0.01%	100% ± 0%	100% ± 0%	100% ± 0%	100% ± 0%
	(UB) LOG LIKELYHOOD	**0.1% ± 0%**	1.39% ± 1.73%	1.14% ± [illegible].85%	5.18% ± 2.09%	100% ± 0%	100% ± 0%	100% ± 0%	100% ± 0%
	(UB) PEARSON WEIGHTED	**0.1% ± 0.01%**	0.71% ± 0.2%	[illegible]65% [illegible]1%	3.41% ± 1.98%	100% ± 0%	100% ± 0%	100% ± 0%	100% ± 0%

Figure 18: Information Leak - Utility Function related information

and the recommendation aggregation one, TRADGRec-RA). Logically, when using *MAGReS [One-Step, M19]*, *MAGReS [MCP Next; M11]* and *MAGReS [MCP Relaxed; M17]* the amount of information leaked is always lower than when using the *MAGReS [MCP Strict; M5]*, as *MAGReS [One-Step; M19]* only uses one negotiation round and both *MAGReS [MCP Next; M11]* and *MAGReS [MCP Relaxed; M17]* reduce the amount negotiation rounds by "making" the agents reach agreements faster (see Sections 3.5 and 4.5.3). The same conclusion was reached when analyzing the amount of information leaked regarding the candidate proposals. We found out that, again, *MAGReS [One-Step; M19], MAGReS [MCP Next; M11]* and *MAGReS [MCP Relaxed; M17]* variants always leak less information than the *MAGReS [MCP Strict; M5]* one. All in all, the best variant out of the four tested seems to be *MAGReS [MCP Relaxed; M17]*, as it leaks a reasonable amount of information while achieving, in many cases, the highest GS, fairness and amount of effective recommendations, and the lowest MSD (see Section 4.5.3).

To confirm the validity of the results, we run statistical tests following the same strategy as the one used in previous sections. For each dataset and for each similarity metric we first performed a normality test and then, when we confirmed that at least one of the samples did not follow the normal distribution, we performed a pair-wise comparison among the samples (5 in Figure 18 and [illegible] in Figure 19) using the Wilcoxon Signed Ranks test. As one approach, A, is better than another one, B, if it leaks less information with regard to the utility

INFORMATION LEAK – PROPOSALS (%) (lower is better)		DATASET: ML_LATEST_SMALL			
		MAGReS [One-Step; M19]	MAGReS [MCP Strict; M5]	MAGReS [MCP Relaxed; M17]	MAGReS [MCP Next; M11]
SIMILARITY METRIC	(IB) CITY BLOCK	**0.01% ± 0%**	0.32% ± 0.37%	0.02% ± 0.02%	0.64% [illegible] 0.6[illegible]
	(IB) EUCLIDEAN WEIGHTED	**0.01% ± 0%**	0.4% ± 0.16%	0.28% ± 0.15%	0.84% ± 0.43%
	(IB) PEARSON WEIGHTED	**0.01% ± 0%**	0.35% ± 0.17%	0.28% ± 0.16%	[illegible] ± 0.39%
	(IB) UNCENTERED COSINE WEIGHTED	**0.01% ± 0%**	0.33% ± 0.14%	0.23% ± 0.11%	0.59% ± 0.34%
	(UB) LOG LIKELYHOOD	**0.01% ± 0%**	0.09% ± 0.04%	0.02% ± 0.0[illegible]	[illegible]% ± 0.0[illegible]
	(UB) PEARSON WEIGHTED	**0.01% ± 0%**	0.28% ± 0.11%	[illegible]1% ± [illegible].08%	[illegible]% ± 0.2[illegible]

INFORMATION LEAK – PROPOSALS (%) (lower is better)		DATASET: [illegible]ELP			
		MAGReS [One-Step; M19]	MAGReS [MCP Strict; M5]	MAGReS [M[illegible]laxed; M1[illegible]]	[illegible]GReS [MCP Next; M11]
SIMILARITY METRIC	(IB) CITY BLOCK	**0% ± 0%**	0.01% ± 0.[illegible]	0% ± 0.01%	0.02% ± 0.02%
	(IB) EUCLIDEAN WEIGHTED	**0% ± 0%**	[illegible].02% ± [illegible]	[illegible].01% [illegible]	0.01% ± 0.01%
	(IB) PEARSON WEIGHTED	**0% ± 0%**	0.01% ± 0.01%	0.01% ± [illegible]01%	0.01% ± 0.01%
	(IB) UNCENTERED COSINE WEIGHTED	**0% ± 0%**	0.02% ± [illegible].01%	0.01% ± 0.01%	11.12% ± 3.58%
	(UB) LOG LIKELYHOOD	**0% ±**	**0% [illegible] 0%**	**0% ± 0%**	**0% ± 0%**
	(UB) PEARSON WEIGHTED	**0[illegible] 0%**	**0[illegible] 0%**	**0% ± 0%**	**0% ± 0%**

Figure 19: Information Leak: Candidate proposals related information

function (Figure 18) and the candidate proposals of each agent (Figure 19), for each test we defined the null hypothesis as "the sample of the variant A is not worse than the sample of the variant B". The results of the test confirmed our findings as, in each test, the null hypothesis was rejected at a significance level of 95% ($\alpha = 0.05$). At last, we confirmed that MAGReS leaked less information than TRADGRec-PA and TRADGRec-PA with respect to the users' utility function, and that, with the exception of *MAGReS [One-Step; M19]*, *MAGReS [MCP; M17]* was also the parameterization that fewer proposals revealed.

4.5.5. Smaller size recommendations (K=3 and K=5)

Once we determined that MAGReS was capable of producing better recommendations than TRADGRec-RA and TRADGRec-PA when the amount of items to be recommended was 10, we wanted to know if the same situation would happen when making smaller recommendations (i.e., with less items). For this matter, we replicated the tests performed for $k = 10$ (see Section 4.4) but just using the similarity metrics that provided the best results in terms of the quality of the recommendations produced: *Euclidean Distance* similarity for the item-based SUR and *Pearson Similarity* for the user-based SUR. The results of the experiments conducted showed that for both $k = 3$ and $k = 5$ MAGReS outperformed the traditional approaches (TRADGRec-PA and TRADGRec-RA) in terms of:

- The GS and the MSD of her recommendations. As it can be seen in Figure 20 and Figure 21 the recommendations produced by all the variants of MAGReS not only achieved a higher level of satisfaction for the group (i.e., the GS) but also were able to satisfy all of its members more uniformly (i.e., the MSD was lower) and increased the fairness of the recommendations.

- The amount of information leaked. As it can be seen in Figure 22, in the worst case scenario (when using *Taboo* as the *ARP* strategy and *Strict* as the *PrA* strategy) MAGReS leaked less utility-function related information than TRADGRec-PA and TRADGRec-PA. Regarding the leak of proposals related information we found out that, as it happened when we set $k = 10$ (see Section 4.5.4), in the worst case scenario the MCP-based variant of MAGReS leaked no more than the 0.35% (for the experiments with ML_LATEST_SMALL)/0.01% (for the experiments with YELP) of the information, which means that over all the items present in the datasets (see Section 4.4), just the 0.35% (ML_LATEST_SMALL) and 0.01% (YELP) of them was effectively proposed by every agent during each one the recommendations. Note that the amount of information (of all types) leaked was significantly lower than when making recommendations of 10 items, and this is explained because a lower amount of items to be recommended implies a lower amount of negotiation processes to be carried out (see Section 4.4), which leads to less information leaked by the agents.

With regard to the analysis of the *ARP* and *PrA* strategies, the tests proved that the observations made in the analysis for $k = 10$ were also valid for $k = 3$ and $k = 5$:

- The use of the *ARP* strategies reduced significantly the amount of "already rated items" being recommended. The *Taboo* variant was able to completely eliminate the overlap for both of the similarity metrics, while not producing a significant negative impact in the quality of the recommendation.

- The *PrA* strategies helped to increase the group satisfaction while also increasing the amount of effective recommendations,reducing the amount of concessions needed to reach the end of the negotiation (either with an agreement or a conflict), increasing the fairness of the recommendations and in some cases, reducing the MSD.

4.5.6. Summary of results

The *ARP* strategy was created in order to model, as an agent-like behavior, how the user feels about receiving a recommendation with an item she/he has already rated. The *ARP* strategy works as a penalty to the utility reported by the agent when asked about a certain proposal. Each variant of the *ARP* strategy has its own rules for computing the penalty. From all the variants of the *ARP* strategy, our tests showed that, independently the dataset used, *Taboo* was the most effective variant at the task of reducing (to zero in most of the cases) the overlap between items recommended and items already rated by the group members, but this came at the cost of reducing

GROUP SATISFACTION (higher is better)		MAGReS [One-Step; M19]	MAGReS [MCP Strict; M5]	MAGReS [MCP Relaxed; M17]	MAGReS [MCP Next; M11]	TRADGRec-PA [AVG; T1]	TRADGRec-PA [UL; T5]	TRADGRec-RA [AVG; T1]	TRADGRec-RA [UL; T5]
DATASET: ML_LATEST_SMALL									
SIMILARITY METRIC	(IB) EUCLIDEAN WEIGHTED	0.8028 ± 0.0835	0.8407 ± 0.1352	**0.8577 ± 0.0424**	0.8305 ± 0.0687	[illegible] ± 0.1727	[illegible] ± 0.1856	0.701 ± 0.106	0.701 ± 0.106
	(UB) PEARSON WEIGHTED	0.8182 ± 0.1192	0.962 ± 0.0264	**0.9638 ± 0.0188**	0.8837 ± 0.0712	0.5396 ± [illegible]	[illegible]5429 ± [illegible].1651	0.7054 ± 0.1537	0.7054 ± 0.1537

GROUP SATISFACTION (higher is better)		MAGReS [One-Step; M19]	MAGReS [MCP Strict; M5]	MAGReS [MCP Relaxed; M17]	MAGReS [MCP Next; M11]	TRADGRec-PA [AVG; T1]	TRADGRec-PA [UL; T5]	TRADGRec-RA [AVG; T1]	TRADGRec-RA [UL; T5]
DATASET: YELP									
SIMILARITY METRIC	(IB) EUCLIDEAN WEIGHTED	0.7105 ± 0.1596	0.7069 ± 0.408	**0.9222 ± 0.042**	0.8894 ± [illegible]617	0.20[illegible] ± [illegible]323	0.1959 ± 0.138	0.5341 ± 0.1459	0.5341 ± 0.1459
	(UB) PEARSON WEIGHTED	0.8644 ± 0.1532	0.9926 ± 0.0497	**0.9902 ± 0.0495**	0.9619 ± [illegible]	0.6745 ± 0.1772	0.7012 ± 0.1815	0.7871 ± 0.1229	0.7871 ± 0.1229

(a) Average GS per similarity metric and approach

MEMBERS SATISFACTION DISPERSION (lower is better)		MAGReS [One-Step; M19]	MAGReS [MCP Strict; M5]	MAGReS [MCP Relaxed; M17]	MAGReS [MCP Next; M11]	TRADGRec-PA [AVG; T1]	TRADGRec-PA [UL; T5]	TRADGRec-RA [AVG; T1]	TRADGRec-RA [UL; T5]
DATASET: ML_LATEST_SMALL									
SIMILARITY METRIC	(IB) EUCLIDEAN WEIGHTED	0.1233 ± 0.0832	**0.078 ± 0.0366**	[illegible]0852 ± 0.[illegible]77	0.1168 ± 0.0802	0.3181 ± 0.1363	0.2827 ± 0.1349	0.2313 ± 0.1053	0.2313 ± 0.1053
	(UB) PEARSON WEIGHTED	0.1373 ± 0.1079	0.02[illegible] ± 0.0238	**[illegible] ± 0.0103**	0.1321 ± 0.094	0.2933 ± 0.1238	0.3051 ± 0.1323	0.2377 ± 0.1199	0.2377 ± 0.1199

MEMBERS SATISFACTION DISPERSION (lower is better)		MAGReS [One-Step; M19]	MAGReS [MCP Strict; M5]	MAGReS [MCP Relaxed; M17]	MAGReS [MCP Next; M11]	TRADGRec-PA [AVG; T1]	TRADGRec-PA [UL; T5]	TRADGRec-RA [AVG; T1]	TRADGRec-RA [UL; T5]
DATASET: YELP									
SIMILARITY METRIC	(IB) EUCLIDEAN WEIGHTED	0.208[illegible] ± 0.0935	**0.0397 ± 0.0347**	0.0593 ± 0.0286	0.0987 ± 0.0706	0.3244 ± 0.2079	0.3142 ± 0.1884	0.2853 ± 0.1265	0.2853 ± 0.1265
	(UB) PEARSON WEIGHTED	0.[illegible]54 ± [illegible].1429	**[illegible] ± 0.0861**	0.0161 ± 0.0857	0.0572 ± 0.1183	0.2426 ± 0.127	0.2036 ± 0.1097	0.2073 ± 0.1123	0.2073 ± 0.1123

(b) Average MSD per similarity metric and approach

FAIRNESS (higher is better)		MAGReS [One-Step; M19]	MAGReS [MCP Strict; M5]	MAGReS [MCP Relaxed; M17]	MAGReS [MCP Next; M11]	TRADGRec-PA [AVG; T1]	TRADGRec-PA [UL; T5]	TRADGRec-RA [AVG; T1]	TRADGRec-RA [UL; T5]
DATASET: ML_LATEST_SMALL									
SIMILARITY METRIC	(IB) EUCLIDEAN WEIGHTED	0.8401 ± 0.1651	**0.9286 ± 0.1629**	0.9279 ± 0.0948	0.8863 ± 0.1103	0.6044 ± 0.212	0.6057 ± 0.2223	0.6885 ± 0.16	0.6885 ± 0.16
	(UB) PEARSON WEIGHTED	0.8575 ± 0.1403	0.9985 ± 0.0099	**1 ± 0**	0.911 ± 0.0849	0.5438 ± 0.167	0.5464 ± 0.1654	0.7186 ± 0.1595	0.7186 ± 0.1595

FAIRNESS (higher is better)		MAGReS [One-Step; M19]	MAGReS [MCP Strict; M5]	MAGReS [MCP Relaxed; M17]	MAGReS [MCP Next; M11]	TRADGRec-PA [AVG; T1]	TRADGRec-PA [UL; T5]	TRADGRec-RA [AVG; T1]	TRADGRec-RA [UL; T5]
DATASET: YELP									
SIMILARITY METRIC	(IB) EUCLIDEAN WEIGHTED	0.7362 ± 0.1887	0.7511 ± 0.4331	**0.9867 ± 0.0505**	0.9391 ± 0.072	0.2044 ± 0.1339	0.2054 ± 0.1443	0.5426 ± 0.1693	0.5426 ± 0.1693
	(UB) PEARSON WEIGHTED	0.8647 ± 0.1583	0.9926 ± 0.0497	**0.9926 ± 0.0497**	0.964 ± 0.0821	0.6746 ± 0.1808	0.7104 ± 0.1896	0.779 ± 0.1319	0.779 ± 0.1319

(c) Average fairness per similarity metric and approach

Figure 20: Average GS, MSD and fairness for recommendations of size 3 ($k = 3$)

GROUP SATISFACTION (higher is better) — DATASET: ML_LATEST_SMALL	MAGReS [One-Step; M19]	MAGReS [MCP Strict; M5]	MAGReS [MCP Relaxed; M17]	MAGReS [MCP Next; M11]	TRADGRec-PA [AVG; T1]	TRADGRec-PA [UL; T5]	TRADGRec-RA [AVG; T1]	TRADGRec-RA [UL; T5]
SIMILARITY METRIC: (IB) EUCLIDEAN WEIGHTED	0.7984 ± 0.076	0.8353 ± 0.1347	**0.8536 ± 0.0425**	0.824 ± 0.0736	[illegible] ± 0.1724	[illegible] ± 0.1684	0.6871 ± 0.115	0.6871 ± 0.115
SIMILARITY METRIC: (UB) PEARSON WEIGHTED	0.78 ± 0.1078	0.949 ± 0.0304	**0.9546 ± 0.0196**	0.8881 ± 0.0539	0.5407 ± [illegible]	[illegible] ± 0.117	0.6838 ± 0.0944	0.6838 ± 0.0944

GROUP SATISFACTION (higher is better) — DATASET: YELP	MAGReS [One-Step; M19]	MAGReS [MCP Strict; M5]	MAGReS [MCP Relaxed; M17]	MAGReS [MCP Next; M11]	TRADGRec-PA [AVG; T1]	TRADGRec-PA [UL; T5]	TRADGRec-RA [AVG; T1]	TRADGRec-RA [UL; T5]
SIMILARITY METRIC: (IB) EUCLIDEAN WEIGHTED	0.6923 ± 0.1345	0.8401 ± 0.2679	**0.918 ± 0.0423**	0.8782 ± [illegible]	[illegible] ± [illegible]	0.1976 ± 0.1258	0.594 ± 0.1215	0.594 ± 0.1215
SIMILARITY METRIC: (UB) PEARSON WEIGHTED	0.8563 ± 0.1134	**0.9926 ± 0.0497**	0.9906 ± 0.0495	0.9661 ± [illegible]	0.6933 ± 0.1401	0.7142 ± 0.1293	0.75 ± 0.1134	0.75 ± 0.1134

(a) Average GS per similarity metric and approach

MEMBERS SATISFACTION DISPERSION (lower is better) — DATASET: ML_LATEST_SMALL	MAGReS [One-Step; M19]	MAGReS [MCP Strict; M5]	MAGReS [MCP Relaxed; M17]	MAGReS [MCP Next; M11]	TRADGRec-PA [AVG; T1]	TRADGRec-PA [UL; T5]	TRADGRec-RA [AVG; T1]	TRADGRec-RA [UL; T5]
SIMILARITY METRIC: (IB) EUCLIDEAN WEIGHTED	0.1194 ± 0.0703	**0.0768 ± 0.0383**	[illegible]	0.1192 ± 0.0861	0.2995 ± 0.1354	0.2709 ± 0.1301	0.219 ± 0.1156	0.219 ± 0.1156
SIMILARITY METRIC: (UB) PEARSON WEIGHTED	0.1233 ± 0.0776	[illegible] ± 0.028	**[illegible] ± 0.0119**	0.0983 ± 0.0589	0.2646 ± 0.0912	0.2622 ± 0.0948	0.2085 ± 0.076	0.2085 ± 0.076

MEMBERS SATISFACTION DISPERSION (lower is better) — DATASET: YELP	MAGReS [One-Step; M19]	MAGReS [MCP Strict; M5]	MAGReS [MCP Relaxed; M17]	MAGReS [MCP Next; M11]	TRADGRec-PA [AVG; T1]	TRADGRec-PA [UL; T5]	TRADGRec-RA [AVG; T1]	TRADGRec-RA [UL; T5]
SIMILARITY METRIC: (IB) EUCLIDEAN WEIGHTED	[illegible] ± 0.0808	**0.0548 ± 0.0315**	0.0563 ± 0.0263	0.0933 ± 0.0725	0.3081 ± 0.1803	0.3079 ± 0.1686	0.2183 ± 0.0809	0.2183 ± 0.0809
SIMILARITY METRIC: (UB) PEARSON WEIGHTED	[illegible] ± [illegible]	**[illegible] ± 0.0861**	0.0154 ± 0.0857	0.0458 ± 0.0946	0.1968 ± 0.1026	0.1714 ± 0.0912	0.1955 ± 0.097	0.1955 ± 0.097

(b) Average MSD per similarity metric and approach

FAIRNES (higher is better) — DATASET: ML_LATEST_SMALL	MAGReS [One-Step; M19]	MAGReS [MCP Strict; M5]	MAGReS [MCP Relaxed; M17]	MAGReS [MCP Next; M11]	TRADGRec-PA [AVG; T1]	TRADGRec-PA [UL; T5]	TRADGRec-RA [AVG; T1]	TRADGRec-RA [UL; T5]
SIMILARITY METRIC: (IB) EUCLIDEAN WEIGHTED	0.8333 ± 0.1514	0.8353 ± 0.1347	**0.9232 ± 0.0947**	0.8886 ± 0.1114	0.5968 ± 0.203	0.615 ± 0.1956	0.6693 ± 0.1649	0.6693 ± 0.1649
SIMILARITY METRIC: (UB) PEARSON WEIGHTED	0.8047 ± 0.1226	0.9947 ± 0.0202	**0.9991 ± 0.006**	0.9174 ± 0.0703	0.4924 ± 0.0909	0.5025 ± 0.0905	0.6921 ± 0.105	0.6921 ± 0.105

FAIRNESS (higher is better) — DATASET: YELP	MAGReS [One-Step; M19]	MAGReS [MCP Strict; M5]	MAGReS [MCP Relaxed; M17]	MAGReS [MCP Next; M11]	TRADGRec-PA [AVG; T1]	TRADGRec-PA [UL; T5]	TRADGRec-RA [AVG; T1]	TRADGRec-RA [UL; T5]
SIMILARITY METRIC: (IB) EUCLIDEAN WEIGHTED	0.7169 ± 0.1712	0.9053 ± 0.2873	**0.985 ± 0.0482**	0.9287 ± 0.0764	0.2002 ± 0.1224	0.2085 ± 0.1305	0.5995 ± 0.1398	0.5995 ± 0.1398
SIMILARITY METRIC: (UB) PEARSON WEIGHTED	0.8619 ± 0.1189	0.9926 ± 0.0497	0.9926 ± 0.0497	**0.9666 ± 0.0861**	0.6967 ± 0.1424	0.7246 ± 0.1336	0.7511 ± 0.1182	0.7511 ± 0.1182

(c) Average fairness per similarity metric and approach

Figure 21: Average GS, MSD and fairness for recommendations of size 5 ($k = 5$)

INFORMATION LEAK - UTILITY FUNCTION (%) (lower is better)		DATASET: ML_LATEST_SMALL							
		MAGReS [One-Step; M19]	MAGReS [MCP Strict; M5]	MAGReS [MCP Relaxed; M17]	MAGReS [[illegible] Next; M1[illegible]]	[illegible]GRec-PA [AVG; T1]	TRADGRec-PA [UL; T5]	TRADGRec-RA [AVG; T1]	TRADGRec-RA [UL; T5]
SIMILARITY METRIC	(IB) EUCLIDEAN WEIGHTED	**2.75% ± 0.01%**	41.04% ± 14.36%	27.97% ± 14.36[illegible]	[illegible] ± 11.48%	100% ± 0%	100% ± 0%	100% ± 0%	100% ± 0%
	(UB) PEARSON WEIGHTED	**2.74% ± 0.01%**	20.46% ± 6.24%	9.93% ± 4.65%	[illegible].17% ± 3.[illegible]	100% ± 0%	100% ± 0%	100% ± 0%	100% ± 0%

INFORMATION LEAK - UTILITY FUNCTION (%) (lower is better)		DATASET: YELP							
		MAGReS [One-Step; M19]	MAGReS [MCP Strict; M5]	MAG[illegible]S [MCP Rela[illegible]; M17]	MAGReS [MCP Next; M11]	TRADGRec-PA [AVG; T1]	TRADGRec-PA [UL; T5]	TRADGRec-RA [AVG; T1]	TRADGRec-RA [UL; T5]
SIMILARITY METRIC	(IB) EUCLIDEAN WEIGHTED	**0.33% ± 0%**	5.6% ± 4.82%	5.[illegible] ± [illegible].22%	6.76% ± 2.23%	100% ± 0%	100% ± 0%	100% ± 0%	100% ± 0%
	(UB) PEARSON WEIGHTED	**0.33% ± 0.02%**	0[illegible] ± 0.2[illegible]	0.6[illegible] ± [illegible]	2.46% ± 1.05%	100% ± 0%	100% ± 0%	100% ± 0%	100% ± 0%

(a) Utility Function related information

INFORMATION LEAK – PROPOSALS (%) (lower is better)		DATASET: ML_LATEST_SMALL			
		MAGReS [One-Step; M19]	MAGReS [MCP Strict; M5]	MAGReS [MCP Relaxed; M17]	MAGReS [MCP Next; M11]
SIMILARITY METRIC	(IB) EUCLIDEAN WEIGHTED	**0.01% ± 0%**	0.33% ± 0.13%	0.2% ± 0.11%	0.75% ± 0.41%
	(UB) PEARSON WEIGHTED	**0.01% ± 0%**	0.16% ± 0.08%	0.06% ± 0.05%	0.28% ± 0.19%

INFORMATION LEAK – PROPOSALS (%) (lower is better)		DATASET: YELP			
		MAGReS [One-Step; M19]	MAGReS [MCP Strict; M5]	MAGReS [MCP Relaxed; M17]	MAGReS [MCP Next; M11]
SIMILARITY METRIC	(IB) EUCLIDEAN WEIGHTED	**0% ± 0%**	0.01% ± 0.01%	0.01% ± 0.01%	0.01% ± 0.01%
	(UB) PEARSON WEIGHTED	**0% ± 0%**	**0% ± 0%**	**0% ± 0%**	**0% ± 0%**

(b) Proposals related information

Figure 22: Information leak for recommendations of size 3 ($k = 3$)

the group satisfaction. The *Flexible* variant, on the other side, provided a good $reduction\,of\,overlap/group\,satisfaction\,loss$ ratio, particularly when the $flexibilityLevel\,(f)$ parameter was set to 0.75. In this case, the tests showed that, for both datasets, the overlap was reduced to almost a 1% overall except in the case of the similarity metric *(IB) City Block*, for which the overlap was around 11% in the experiments performed with ML_LATEST_SMALL and 2% in those performed with YELP. Additionally, the group satisfaction loss was generally lower than when using the *Taboo* variant. With regard to the fairness, the *Min-Satisfaction* (for ML_LATEST_SMALL) and *Easy-Going* (for YELP) variants seem to perform the best but, due to the how these variants work, the overlap is high. All in all, depending on which ARF strategy is selected by the group members for their agents, the group recommendation will change.

The *PrA* strategy was introduced as a way to improve the model of users criteria with regards to the decision of accepting a proposal. Three different variants were proposed, namely: *Strict* (originally proposed in [12]), *Relaxed* and *Next*. For our tests we defined scenarios where all the agents used the same *PrA* strategy, so as to analyze the effects of each strategy on the recommendations. We observed that, independently the dataset used, when all the agents used the *Strict* variant the amount of effective recommendations was lower than when they used either the *Next* or the *Relaxed* variants. This was caused by the amount of negotiations that ended up with no agreement (i.e., with conflict), as many proposals were rejected because of not being exactly what the agents were expecting. Naturally, when the agents are more flexible when determining whether to accept or not a proposal, they are more prone to accept proposals that are not exactly what they wanted but that are good enough, and so the amount of effective recommendations increases. In terms of the quality of the recommendations measured by the group satisfaction and how uniformly where the group members satisfied, we observed that when the *Relaxed* variant was used there were some improvements (if compared to the *Strict* variant), but when the *Next* variant was used the quality of the recommendations produced was lower (due to the the sub-optimal decisions taken by the agents). Additionally, we observed that the use of the variants *Relaxed* and *Next* helped to improve the fairness of the recommendations regardless of the dataset used.

Finally, regarding to the framework that supports the approach, we have also added support for the use of Mahout-based SURs as they were faster than the Duine-based SURs for both loading the data models and generating the recommendations. The similarity metrics available depended on the type of SUR used. The tests revealed that some of them were not able to produce reliable recommendations (for example, *City Block* and *Euclidean Distance* similarities for the user-based SUR) and therefore have to be treated with care.

The main insights from the experiments were the following:

- The *Relaxed* variant of the PrA strategy can be considered as the "ideal variant", as it not only helps to increase the number of recommended items but also reduces the number of concessions, without negatively impacting on the quality of the recommendations.

- Out of all the variants of the ARP strategy, *Taboo* performed the best at the task of preventing MAGReS from recommending items already rated by the group members. On the downside, this variant might decrease the quality of the recommendations , but overall it is still good enough. With this consideration in mind, we would recommended that all the agents use this strategy by default, unless their users decide to manually change it.In terms of performance, the One-Step protocol allowed MAGReS to make faster recommendations, when compared to the MCP , but this came at the cost of drops in the recommendation quality.

- MCP, in turn, helped to improve the quality of the recommendations by increasing the group satisfaction and also satisfying all the group members more uniformly.

- With regard to the SUR, we noticed that the best recommendations were generated when using the *Euclidean Distance* similarity for the item-based SUR and the *Pearson Similarity* for the user-based SUR.

5. Conclusions and Future Work

In this article we proposed MAGReS, a group recommendation approach based on MAS as an alternative to the traditional approaches, which employ a combination between aggregation techniques and SUR techniques to produce group recommendations. In MAGReS, on the contrary, the agents act on behalf of the users, protecting their interests and representing them in a negotiation process that mimics the way humans negotiate about a certain topic. The results of the experiments showed that the use of negotiation instead of the aggregation techniques can greatly improve the quality of the recommendations, not only increasing the level of satisfaction of the group as a whole (group satisfaction) but also satisfying the group members more uniformly (i.e. by reducing MSD and increasing fairness). Along this line, the inclusion of the *ARP* and *PrA* strategies had an impact on the the recommendations produced and allows the users to personalize the behavior of the agents representing them in the recommendation process.

Although we obtained satisfactory results, our experimental study has some limitations. First, the user groups were selected randomly from the dataset, and so we ignored any potential relationships of group members. However, it might be the case that "similarities" between particular users (e.g., friendship, common tastes, etc.) within the group might affect the recommendations, and thus, change the resulting average satisfaction for some groups [8]. Also, user relationships of trust and influence might affect the item selection. There are approaches for both automatically detecting groups [47, 48] (so as to avoid the random selection of users) and dealing with social relationships among the group members [49], so we plan to tackle these aspects in future work. Second, the reliance of our current implementation of the users' utility function on the prediction made by the SURs. All the rating predictions are made based on the configuration of the SUR and so increasing the quality of the predictions, by using a different approach like the one proposed in [50], can help to improve the recommendations produced by MAGReS. According to our findings, the datasets were small and their rating matrices were no dense enough for the similarity metrics to work properly when making recommendations using a user-based SUR. This factor impacted on how the utility function of the agents works, and thus in the recommendations generated by MAGReS.

As a future work we plan to: (*i*) to compare our approach with other techniques for group recommendation, (*ii*) evaluate our approach with real users, (*iii*) assess the approach in a dataset with a more dense rating matrix, (*iv*) study alternative utility functions for the agents, and (*v*) analyze new variants for both the *ARP* and PrA strategies.

Acknowledgments. This work has been partially supported by CONICET PIP Project Number 112-201501-00030 and ANPCyT PICT Project Number 2016-2973.

References

[1] Paul Rogers, Rudy Puryear, and James Root. Infobesity: The enemy of good decisions, 2013.

[2] Christopher C. Yang, Hsinchun Chen, and Kay Hong. Visualization of large category map for internet browsing. *Decis. Support Syst.*, 35(1):89–102, April 2003.

[3] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez. Recommender systems survey. *Knowledge-Based Systems*, 46:109 – 132, 2013.

[4] Francesco Ricci, Lior Rokach, and Bracha Shapira. *Recommender Systems Handbook*. Springer-Verlag New York, Inc., New York, NY, USA, 2nd edition, 2015.

[5] Anthony Jameson and Barry Smyth. *Recommendation to Groups*, pages 596–627. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[6] Linas Baltrunas, Tadas Makcinskas, and Francesco Ricci. Group recommendations with rank aggregation and collaborative filtering. In *Proc. of RecSys '10*, pages 119–126. ACM, 2010.

[7] Ivan Cantador and Pablo Castells. Group recommender systems: New perspectives in the social web. In *Recommender Systems for the Social Web*, volume 32 of *Intelligent Systems Reference Library*, pages 139–157. Springer Berlin Heidelberg, 2012.

[8] Judith Masthoff. *Recommender Systems Handbook*, chapter Group Recommender Systems: Combining Individual Models, pages 677–702. Springer Science+Business Media, 2011.

[9] J. Castro, J. Lu, G. Zhang, Y. Dong, and L. MartÃnez. Opinion dynamics-based group recommender systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pages 1–13, 2017.

[10] Thuy Ngoc Nguyen and Francesco Ricci. Dynamic elicitation of user preferences in a chat-based group recommender system. In *Proceedings of the Symposium on Applied Computing*, SAC '17, pages 1685–1692. ACM, 2017.

[11] Tuomas W. Sandholm. Multiagent systems. chapter Distributed Rational Decision Making, pages 201–258. MIT Press, Cambridge, MA, USA, 1999.

[12] Christian Villavicencio, Silvia Schiaffino, J. Andres Diaz-Pace, Ariel Monteserin, Yves Demazeau, and Carole Adam. *A MAS Approach for Group Recommendation Based on Negotiation Techniques*, pages 219–231. Springer International Publishing, Cham, 2016.

[13] Ulle Endriss. Monotonic concession protocols for multilateral negotiation. In *Proc. AAMAS 2006*, pages 392–399. New York, NY, USA, 2006. ACM.

[14] Michael Wooldridge. *An Introduction to MultiAgent Systems*. John Wiley & Sons, second edition edition, 2009.

[15] Wei Wang, Guangquan Zhang, and Jie Lu. Member contribution-based group recommender system. *Decision Support Systems*, 87:80 – 93, 2016.

[16] Mark O'Connor, Dan Cosley, Joseph A. Konstan, and John Riedl. Polylens: a recommender system for groups of users. In *Proc. of ECSCW'01*, pages 199–218. Kluwer Academic Publishers, 2001.

[17] Wei Wang, Guangquan Zhang, and Jie Lu. Hierarchy visualization for group recommender systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2017.

[18] Ingrid Christensen and Silvia Schiaffino. A hybrid approach for group profiling in recommender systems. *J. of Universal Computer Science*, 20(4):507–533, 2014.

[19] Judith Masthoff. Group modeling: Selecting a sequence of television items to suit a group of viewers. *User Modeling and User-Adapted Interaction*, 14(1):37–85, 2004.

[20] Judith Masthoff. Group recommender systems: Aggregation, satisfaction and group attributes. In *Recommender Systems Handbook*, 2015.

[21] P. Bekkerman, K. Sarit, and F. Ricci. Applying cooperative negotiation methodology to group recommendation problem. In *ECAI Workshop on Recommender systems*, 2006.

[22] Joao Sousa Lopes, Sergio Alvarez-Napagao, Roberto Confalonieri, and Javier Vázquez-Salceda. Use: a multi-agent user-driven recommendation system for semantic knowledge extraction. 2009.

[23] Domenico Rosaci and Giuseppe M. L. Sarné. A multi-agent recommender system for supporting device adaptivity in e-commerce. *J Intell Inf Syst*, 38(2):393–418, may 2011.

[24] Punam Bedi, Sumit Kumar Agarwal, Vinita Jindal, and Richa. Marst: Multi-agent recommender system for e-tourism using reputation based collaborative filtering. In *Proceedings of the 9th International Workshop on Databases in Networked Information Systems - Volume 8381*, DNIS 2014, pages 189–201, New York, NY, USA, 2014. Springer-Verlag New York, Inc.

[25] Inma Garcia, Laura Sebastia, and Eva Onaindia. A negotiation approach for group recommendation. In *Proceedings of the 2009 International Conference on Artificial Intelligence,*, pages 919–925, 2009.

[26] Fabiana Lorenzi, Ana Bazzan, and Mara Abel. An Architecture for a Multiagent Recommender System in Travel Recommendation Scenarios. In *ECAI Workshop on Recommender Systems*, volume 1, pages 88–91, 2006.

[27] V. N. Marivate, G. Ssali, and T. Marwala. An intelligent multi-agent recommender system for human capacity building. In *MELECON 2008 - The 14th IEEE Mediterranean Electrotechnical Conference*, pages 909–915, May 2008.

[28] Jorge Morais, Eugenio Oliveira, and Alipio Jorge. *Distributed Computing and Artificial Intelligence*, volume 151 of *Advances in Intelligent and Soft Computing*, chapter A Multi Agent Recommender System, pages 281–288. Springer, 2012.

[29] Paula Rodríguez, Néstor Duque, and Demetrio A. Ovalle. *Multi-agent System for Knowledge-Based Recommendation of Learning Objects Using Metadata Clustering*, pages 356–364. Springer International Publishing, Cham, 2015.

[30] Debbie Zhang, Simeon Simoff, Silvana Aciar, and John Debenham. A multi agent recommender system that utilises consumer reviews in its recommendations. *International Journal of Intelligent Information and Database Systems*, 2(1):69, 2008.

[31] Yolanda Blanco-Fernández, José J. Pazos-Arias, Alberto Gil-Solla, Manuel Ramos-Cabrer, Belén Barragáns-Martínez, Martín López-Nores, Jorge García-Duque, Ana Fernández-Vilas, and Rebeca P. Díaz-Redondo. *AVATAR: An Advanced Multi-agent Recommender System of Personalized TV Contents by Semantic Reasoning*, pages 415–421. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[32] Wei-Po Lee. Towards agent-based decision making in the electronic marketplace: interactive recommendation and automated negotiation. *Expert Systems with Applications*, 27(4):665 – 679, 2004.

[33] L. Sebastiá, As. Giret, and I. García. A multi agent architecture for single user and group recommendation in the tourism domain. 2011.

[34] Pavle Skocir, Luka Marusic, Marinko Marusic, and Ana Petric. *Agent and Multi-Agent Systems. Technologies and Applications*, volume 7327 of *LNCS*, chapter The MARS - A Multi-Agent Recommendation System for Games on Mobile Phones, pages 104–113. Springer, 2012.

[35] Inma Garcia and Laura Sebastia. A negotiation framework for heterogeneous group recommendation. *Expert Systems with Applications*, 41(4,1):1245 – 1261, 2014.

[36] Silvia Rossi, Claudia Di Napoli, Francesco Barile, and Luca Liguori. Conflict resolution profiles and agent negotiation for group recommendations. 2016.

[37] Francesco Barile, Antonio Caso, and Silvia Rossi. Group recommendation for smart applications: A multi-agent view of the problem. In *WOA*, volume 1260 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2014.

[38] L. C. Lee, H. S. Nwana, D. T. Ndumu, and P. De Wilde. The stability, scalability and performance of multi-agent systems. *BT Technology Journal*, 16(3):94–103, 1998.

[39] F. L. B. Zeuthen. *Problems of Monopoly and Economic Warfare*. Routledge and Sons, London, UK, 1930.

[40] Christian Villavicencio, Silvia Schiaffino, J. Andres Diaz-Pace, and Ariel Monteserin. Using negotiation for group recommendation: a user-study on the movies domain. In *SADIO Electronic Journal (EJS)*, volume 17, 2018.

[41] Shaheen S. Fatima, Michael Wooldridge, and Nicholas R. Jennings. An agenda-based framework for multi-issue negotiation. *Artificial Intelligence*, 152(1):1 – 45, 2004.

[42] Alexander Felfernig, Martin Stettinger, Ludovico Boratto, and Marko Tkalcic. *Group Recommender Systems: An Introduction*. Springer Briefs in Electrical and Computer Engineering. Springer US, 3 2018.

[43] Jeffrey S. Rosenschein and Gilad Zlotkin. *Rules of Encounter: Designing Conventions for Automated Negotiation Among Computers*. MIT Press, Cambridge, MA, USA, 1994.

[44] Christian Villavicencio, Silvia Schiaffino, Jorge Andres Diaz-Pace, and Ariel Monteserin. *PUMAS-GR: A Negotiation-Based Group Recommendation System for Movies*, pages 294–298. Springer International Publishing, Cham, 2016.

[45] C. Trattner, A. Said, L. Boratto, and A. Felfernig. *Evaluating Group Recommender Systems*, chapter 3, page 14. Springer, 2018.

[46] Young Duk Seo, Young Gab Kim, Euijong Lee, Kwang Soo Seol, and Doo Kwon Baik. An enhanced aggregation method considering deviations for a group recommendation. *Expert Systems with Applications*, 93:299–312, 2018.

[47] Ludovico Boratto, Salvatore Carta, Alessandro Chessa, Maurizio Agelli, and M. Laura Clemente. Group recommendation with automatic identification of users communities. In *Proc. of the 2009 IEEE/WIC/ACM Int. Joint Conf. on Web Intelligence and Intelligent Agent Technology*, volume 3, pages 547–550. IEEE, 2009.

[48] Ludovico Boratto and Salvatore Carta. Art: group recommendation approaches for automatically detected groups. *International Journal of Machine Learning and Cybernetics*, 6(6):953–980, 2015.

[49] Ingrid Alina Christensen and Silvia Schiaffino. Social influence in group recommender systems. *Online Information Review*, 38(4):524–542, 2014.

[50] L. Boratto, S. Carta, G. Fenu, F. Mulas, and P. Pilloni. Influence of rating prediction on group recommendation accuracy. *IEEE Intelligent Systems*, 31(6):22–27, Nov 2016.

Appendix A. Experimental Parameters Summary

In this appendix we will sum up all the parameters taken into consideration in each of the experiments performed.

Appendix A.1. Similarity metrics

Mahout provides two types of recommendation techniques based on Collaborative Filtering: one user-based and one item-based. As it can be seen in Figure A.23, for each type of technique different similarity metrics are available. For example, whilst *Euclidean Distance* similarity is available for both techniques, *Spearman Correlation* is only available for the user based recommenders and *Adjusted Cosine* can only be used with the item-based recommenders.

Recommender Type	Similarity	Variants	Parameterization ID
ITEM-BASED	Adjusted Cosine	Non-Weighted	N/A
		Weighted	N/A
	City Block	None	(IB) City Block
	Euclidean Distance	Non-Weighted	N/A
		Weighted	(IB) Euclidean Weighted
	Log Likelyhood	None	N/A
	Pearson Correlation	Non-Weighted	N/A
		Weighted	(IB) Pearson Weighted
	Tanimoto Coefficient	None	N/A
	Uncentered Cosine	Non-Weighted	N/A
		None	(IB) Uncentered Cosine Weighted
USER-BASED	City Block	None	(UB) City Block
	Euclidean Distance	Non-Weighted	N/A
		Weighted	(UB) Euclidean Distance Weighted
	Log Likelyhood	None	(UB) Log Likelyhood
	Pearson Correlation	Non-Weighted	N/A
		Weighted	(UB) Pearson Correlation Weighted
	Spearman Correlation	None	N/A
	Uncentered Cosine	Non-Weighted	N/A
		Weighted	(UB) Uncentered Cosine Weighted

Figure A.23: Similarity Metrics supported by each type of SUR (green = used in the experiments, red = tested and discarded, N/A = it was not used in the experiments)

Appendix A.2. Approaches parameters

Appendix A.2.1. TRADGRec-PA and TRADGRec-RA parameters

For TRADGRec-PA and TRADGRec-RA there is only one parameter than can be tuned: the *preference aggregation strategy*. At the moment the approach supports only 5 aggregation strategies, which are specified in Figure A.24. The parameters of those strategies can be found in Figure A.25.

Appendix A.2.2. MAGReS parameters

In MAGReS many parameters can be tuned and each one of them can assume many possible values. In Figure A.26 we specify the most relevant parameters for the experiments carried out in this paper and their possible values. As it

Preference/Rating Aggregation Strategy (PA Strategy)	
Useful for:	To aggregate ratings/preferences
Used by:	TRADGRec-PA, TRADGRec-RA
Name	**Abbreviation + Parameters**
Average (AVG)	AVG
Least Missery (LM)	LM
Most Pleasure (MP)	MP
Approval voting (AV)	AV [approvalVotingThreshold]
Upward Leveling (UL)	UL [alpha,beta, gamma, approvalVotingThreshold]

Figure A.24: TRADGRec preference aggregation strategies

Strategy Parameter	Abbreviation	Value Range	Tested values	Used in comparisons (best results)
approvalVotingThreshold	av_t	[0; 1]	0,2; 0,6; 0,8; 0,9	0,8*
alpha	a	[0; 1]	0,2; 0,4*	0,2*
beta	b	[0; 1]	0,1*	0,1*
gamma	g	[0; 1]	0,5; 0,7*	0,7*

Figure A.25: TRADGRec preference aggregation strategies parameters. *values were extracted from [46]

can be seen in the figure some of the strategies can be further customized with their own parameters. The range of valid values, the values we have used in the tests and the ones we used when doing comparisons among variants of the same strategy are detailed in Figure A.27.

Appendix A.3. Approaches parameterizations

Appendix A.3.1. TRADGRec-PA and TRADGRec-RA parameterizations

In Table A.28 we specify the most relevant parameterizations for TRADGRec-PA and TRADGRec-RA.

Appendix A.3.2. MAGReS parameterizations

Taking into consideration the tables A.26 and A.27 it is possible to see that there is a high amount of possible parameterizations for MAGReS. The Table A.29 details all the parameterizations used to test the approach for this paper, each one of them selected because they were relevant to the experiment we needed to perform. For example, the *One Step* protocol performs only one negotiation round which renders the *PrA* strategies *Next* and *Relaxed* useless. In the first case because it relies on what the agents would do in the next negotiation round, and in the second case because of how the agreement is determined (according to this protocol the agents agree on the item with the higher utility product, which means that the *PrA* strategy is not used).

Each parameterization is identified by an ID which can be used to refer to the parameterization itself. For example, saying that the parameterization *M1* was used means that we set the protocol to *MCP*, the *PrA* strategy to *Strict*, the

Protocol	
Useful for:	Determining how the agents should negotiate
Name	**Abbreviation + Parameters**
Monotonic Concession Protocol	MCP
One-Step Protocol	ONE-STEP

(a) Protocols

Concession Strategy	
Useful for:	Defining the set of rules the agents follow when choosing what to propose in the next round (if they have to make a concession)
Name	**Abbreviation + Parameters**
Desires Distance	DD
Nash	NASH
Utilitarian	UTILITARIAN

(b) Concession Strategies

Already Rated Punishment Strategy (ARP Strategy)	
Useful for:	Model how the users may react when items that she has already rated are proposed to her and make the agent behave in a similar way
Name	**Abbreviation + Parameters**
Easygoing	EASYGOING
Flexible	FLEXIBLE [f=X]
FlexiblePlus [flexibilityLevel, minSatisfaction]	FLEXIBLEPLUS [f=X, ms=Y]
MinSatisfaction [minSatisfaction]	MINSAT [ms=X]
Taboo	TABOO

(c) ARP strategies

Proposal Acceptance Strategy (PrA Strategy)	
Useful for:	Model the way the users determine whether to accept or reject a proposal and make the agent behave in a similar way
Name	**Abbreviation + Parameters**
Strict	STRICT
Next	NEXT
Relaxed	RELAX [rp=X]

(d) PrA strategies

Figure A.26: MAGReS parameters

Strategy Parameter	Abbreviation	Strategy	Value Range	Tested values	Used in comparisons (best results)
relaxPercentage	rp	Relaxed (PrA)	[0; 0,1]	0,025; 0,05; 0,1	0,025; 0,05; 0,1
flexibilityLevel	f	Flexible (ARP)	[0; 1]	0,25; 0,5; 0,75	0.75
minSatisfaction	ms	FlexiblePlus (ARP), MinSatisfaction (ARP)	[0; 1]	0,2; 0,4; 0,6; 0,8	0.6

Figure A.27: MAGReS strategies parameters

TRADGRec-PA	APPROACH PARAMETERS					PARAMETERIZATION
	PA Strategy					
	Name	Parameters				ID
		av_t	a	b	g	
	AVG	N/A				T1
	LM	N/A				T2
	MP	N/A				T3
	AV	0.8	N/A			T4
	UL	0.8	0.2	0.1	0.7	T5

TRADGRec-RA	APPROACH PARAMETERS					PARAMETERIZATION
	PA Strategy					
	Name	Parameters				ID
		av_t	a	b	g	
	AVG	N/A				T1
	LM	N/A				T2
	MP	N/A				T3
	AV	0.8	N/A			T4
	UL	0.8	0.2	0.1	0.7	T5

Figure A.28: TRADGRec-RA and TRADGRec-PA parameterizations

Concession strategy to *Desires Distance* and the *ARP* strategy to *Easy-Going.*

Appendix A.4. Experiments parameterizations

Appendix A.4.1. Single-User-Recommender experiment

For each of the similarity metrics listed with green background in the Table A.23 we analized the results of the experiments that used the following parameterizations:

- For MAGReS:
  - One Step protocol: *M19*
  - MCP protocol: *M5*, *M11*, *M13* and *M17*.
- For TRADGRec-PA: *T1* and T5.
- For TRADGRec-RA: *T1* and T5.

Appendix A.4.2. ARP experiment

To evaluate the impact of the *ARP* strategies on the agents behavior we performed experiments comparing the recommendations produced by the following MAGReS parameterizations: *M1*, *M2*, *M3*, *M4* and *M5*.

Appendix A.4.3. PrA experiment

The MAGReS parameterizations used in the experiments were: $M5$, $M11$, *M13*, *M15* and *M17*.

	APPROACH PARAMETERS				PARAMETERIZATIONS
	PROTOCOL	PrA	CONCESSION STRATEGY	ARP	ID
MAGReS	MCP	STRICT	DD	EASYGOING	M1
				FLEXIBLE [f=0.75]	M2
				FLEXIBLEPLUS [f=0.75, ms=0.6]	M3
				MINSAT [ms=0.6]	M4
				TABOO	M5
			NASH	EASYGOING	M6
				TABOO	M7
			UTILITARIAN	EASYGOING	M8
				TABOO	M9
		NEXT	DD	EASYGOING	M10
				TABOO	M11
		RELAXED [rp=0.025]	DD	EASYGOING	M12
				TABOO	M13
		RELAXED [rp=0.05]	DD	EASYGOING	M14
				TABOO	M15
		RELAXED [rp=0.1]	DD	EASYGOING	M16
				TABOO	M17
	ONE STEP	STRICT	DD	EASYGOING	M18
				TABOO	M19
			NASH	EASYGOING	M20
				TABOO	M21
			UTILITARIAN	EASYGOING	M22
				TABOO	M23

Table A.1: MAGReS parameterizations

Appendix A.4.4. Information privacy experiment

For this experiment we decided to compare those parameterizations that produced the best results in terms of information privacy.

- For TRADGRec-PA: *T1* and T5.
- For TRADGRec-RA: *T1* and T5.
- For MAGReS: *M5*, *M11*, *M17*, *M19*.